

**DESIGNING ENGLISH LISTENING ASSESSMENT FOR THE
FIFTH GRADE STUDENTS OF SDN REJODANI NGAGLIK
SLEMAN YOGYAKARTA**

A THESIS

**Presented as a Partial Fulfilment of the Requirements for the Attainment of
the *Sarjana Pendidikan* Degree in English Language Education**



By:

Amanda Widyati

07202241026

**ENGLISH EDUCATION DEPARTMENT
FACULTY OF LANGUAGES AND ARTS
YOGYAKARTA STATE UNIVERSITY**

2013

APPROVAL SHEET

DESIGNING ENGLISH LISTENING ASSESSMENT FOR THE FIFTH GRADE STUDENTS OF SDN REJODANI NGAGLIK SLEMAN YOGYAKARTA

A THESIS

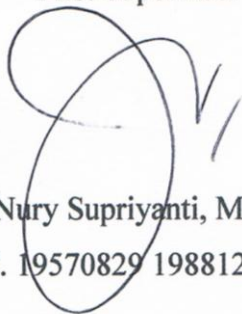
By:

Amanda Widyati

07202241026

Approved on October , 2013

First Supervisor



Nury Supriyanti, M.A.

NIP. 19570829 198812 2 001

Second Supervisor



Lusi Nurhayati, M.Appl. Ling

NIP. 19790205 200312 2 001

RATIFICATION SHEET


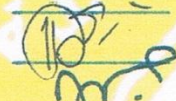


DESIGNING ENGLISH LISTENING ASSESSMENT FOR THE FIFTH GRADE STUDENTS OF SDN REJODANI NGAGLIK SLEMAN YOGYAKARTA

A THESIS


By:
Amanda Widyati
07202241026

Accepted by the Board of Examiners of Faculty of Languages and Arts of State University of Yogyakarta on October 28, 2013 and declared to have fulfilled the requirements for the attainment of a *Sarjana Pendidikan* degree in English Language Education.

Board of Examiners

Position	Name	Signature	Date
Chairperson	: Drs. Suharso, M.P.d.		29-10/13
Secretary	: Lusi Nurhayati, M.Appl. Ling		30-10/13
First Examiner	: Dra. Jamilah, M.Pd.		29-10/13
Second Examiner	: Nury Supriyanti, M.A.		29-10/13

Yogyakarta, October 30, 2013
Faculty of Languages and Arts
State University of Yogyakarta


Zamzani, M. Pd.
NIP. 195505 198011 1 01

PERNYATAAN

Yang bertanda tangan di bawah ini, saya

Nama : **Amanda Widyati**

NIM : 07202241026

Program Studi : Pendidikan Bahasa Inggris

Fakultas : Bahasa dan Seni Universitas Negeri Yogyakarta

menyatakan bahwa karya ilmiah ini adalah hasil pekerjaan saya sendiri. Sepanjang pengetahuan saya, karya ilmiah ini tidak berisi materi yang ditulis oleh orang lain, kecuali bagian-bagian tertentu yang saya ambil sebagai acuan dengan mengikuti tata cara dan etika penulisan karya ilmiah yang lazim.

Apabila ternyata terbukti bahwa pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab saya.

Yogyakarta, Oktober 2013

Penulis,

A handwritten signature in black ink, consisting of stylized, overlapping loops and curves, representing the name Amanda Widyati.

Amanda Widyati

DEDICATIONS

The thesis is dedicated to:

Allah SWT

My parents

My brother and sister

My future



MOTTOS

Allah, tidak ada tuhan selain Dia. Yang Mahahidup, Yang terus-menerus mengurus (makhluk-Nya), tidak mengantuk, dan tidak tidur. Milik-Nya apa yang ada di langit dan apa yang ada di bumi. Tidak ada yang dapat memberi syafaat di sisi-Nya tanpa izin-Nya. Dia mengetahui apa yang di hadapan mereka dan apa yang di belakang mereka, dan mereka tidak mengetahui sesuatu apapun tentang ilmu-Nya melainkan apa yang Dia kehendaki. Kursi-Nya meliputi langit dan bumi. Dan Dia tidak merasa berat memelihara keduanya, dan Dia Mahatinggi, Mahabesar.

(Q.S. Al Baqarah / 2: 255)

Untukmu agamamu, dan untukku agamaku.

(Q.S. Al Kafirun/ 109: 255)

(Allah) tidak beranak dan tidak pula diperanakkan. Dan tidak ada sesuatu yang setara dengan Dia.

(Q.S. Al Iklash/ 112: 255)



ACKNOWLEDGEMENTS

Alhamdulillah hirobbil ‘alamin. Praise is to Allah SWT, the most Gracious and Merciful, without which I would never finish this thesis. I would also like to express my gratitude and appreciation to the following people who have helped, motivated, supported and inspired me during the process of finishing the thesis.

1. My first consultant, Nury Supriyanti, M.A for the precious time, expert guidance and trust given to me. I would like to thank you for helping me to finish my thesis.
2. My second consultant, Lusi Nurhayati, M.Appl. Ling for the support, encouragement and the expert guidance given to me. I would like to thank you for helping me to finish my thesis.
3. Ani Setyaningsih, S. Pd as one of the content and design experts for the advice and guidance.
4. My beloved parents, Bambang and Supriyati. I am so sorry for making you wait so long for my graduation.
5. My beloved brother and sister, Andika and Amara. Thanks for the support and prayer.
6. My boyfriend, Bayu Septian Handayamurti. Thanks for the love, support and for accompanying me since we were in Junior High School.
7. The School Principal of SDN Rejodani, Ponimin, S.Ag, the English teacher of the SDN Rejodani, Dewi Indriyani, S. Pd and the classroom teacher of the fifth grade in SDN Rejodani, Ika Maisari, S.Pd. I thank you for the guidance and suggestions.
8. The fifth grade students of SDN Rejodani for being nice during the try-out.
9. All of my friends in Class B 2007, Weti, Riani, Kaka, Irul, Refri, Ika, Erna, Tama, Adi, Popo, Wulan, Afi, Rosa, Dhanang, Tiya, Windy, Isty and Coco. Thank you for the support and togetherness.
10. My best friends, Erna Sulistyani and Victoria Ika Sartika Dewi. Thanks for the support and never ending friendship.

11. All of the people who always help and support me. I would thank you so much.

Finally, I realize that the thesis is far from being perfect. However, I expect that it gives inspiration and motivation to other people to improve the English assessment for young learners.

Yogyakarta, October 2013

The writer

TABLE OF CONTENTS

TITLE.....	i
APPROVAL SHEET.....	ii
RATIFICATION SHEET.....	iii
<i>PERNYATAAN</i>	iv
DEDICATIONS.....	v
MOTTOS.....	vi
ACKNOWLEDGEMENTS.....	vii
TABLE OF CONTENTS.....	ix
LIST OF FIGURES.....	xvi
LIST OF TABLES.....	xvii
LIST OF APPENDICES.....	xxvi
ABSTRACT.....	xxvii
CHAPTER I INTRODUCTION.....	1
A. Background of the Problem.....	1
B. Identification of the Problem.....	3
C. Limitation of the Problem.....	5
D. Formulation of the Problem.....	6
E. Objective of the Study.....	6
F. Significance of the Study.....	6
CHAPTER II LITERATURE REVIEW.....	7
A. Theoretical Review.....	7
1. Teaching English to Young Learners.....	7
2. Assessment for Young Learners.....	7
a. The Nature of Assessment.....	8
b. Young Learners as the Test-Takers.....	9
1) The Characteristics of Young Language Learners.....	9
2) Issues on Assessing Young Learners.....	11
c. Types of Assessment.....	12
1) Assessment based on the Techniques.....	13

a) Formal Assessment.....	13
b) Informal Assessment.....	13
2) Assessment based on the Time Implementation.....	14
a) Formative Assessment.....	14
b) Summative Assessment.....	14
3) Assessment based on the References.....	14
a) Norm-Referenced Assessment.....	15
b) Criterion-Referenced Assessment.....	15
d. Principles of Language Assessment.....	15
1) Validity.....	15
a) Content Validity.....	16
b) Construct Validity.....	16
c) Face Validity.....	17
2) Reliability.....	18
3) Practicality.....	19
4) Authenticity and Interactiveness.....	19
5) Backwash Effect.....	20
6) Fairness.....	21
e. Cambridge Test for Young Learners as the Standardized Test for Young Language Learners.....	21
1) Listening.....	23
2) Reading and Writing.....	23
3) Speaking.....	24
f. Steps to Design an Assessment.....	25
1) The Designing Stage.....	25
2) The Operationalization Stage.....	27
a) The Test Structure.....	28
b) The Test Task Specifications.....	29
i) The Objective of the Assessment.....	29
ii) Characteristics of the Input.....	29
(1) Input of Format Characteristics.....	29

(2) Input of Language Characteristics.....	30
iii) Characteristics of the Output.....	31
iv) Time Allotment.....	31
c) Criteria of Correctness.....	31
3) The Administration Stage.....	32
3. The Nature of Listening for Young Learners.....	35
a. Teaching Listening to Young Learners.....	35
b. Listening Assessment for Young Learners.....	36
B. Conceptual Framework.....	38
CHAPTER III RESEARCH METHOD.....	42
A. Research Design.....	42
B. Research Subject.....	43
C. Research Setting.....	43
D. Research Instrument.....	43
1. Questionnaires.....	43
2. Observation Sheet.....	46
3. Interviews.....	46
E. Validity and Reliability of the Instrument.....	47
1. Questionnaires.....	47
2. Observation Sheet.....	47
3. Interviews.....	48
F. Data Collection Techniques.....	48
G. Data Analysis Techniques.....	48
H. Research Procedure.....	51
CHAPTER IV RESEARCH FINDINGS AND DISCUSSIONS.....	55
A. Research Findings.....	54
1. Designed Listening Assessment.....	55
a. The Result of Learners' Needs Analysis.....	55
b. Blueprint of the Listening Assessment.....	60
1) The Assessment Structure.....	60
2) The Assessment Task Characteristics.....	63

3) Criteria of Marking.....	69
c. Specifications of the Listening Assessment.....	71
1) First Assessment (<i>In the Classroom</i>).....	71
2) Second Assessment (<i>A Cup of Tea</i>).....	74
3) Third Assessment (<i>Good Daily Activities</i>).....	77
2. Getting Expert Judgement, Evaluating and Revising the First Draft of the Listening Assessment.....	80
a. Getting Expert Judgement of the First Draft of the Listening Assessment.....	80
b. Evaluating the First Draft of the Listening Assessment.....	81
1) First Assessment (<i>In the Classroom</i>).....	81
2) Second Assessment (<i>A Cup of Tea</i>).....	86
3) Third Assessment (<i>Good Daily Activities</i>).....	90
c. The Revision of the First Draft of the Listening Assessment.....	94
3. Trying Out, Evaluating and Revising the Second Draft of the Listening Assessment.....	95
a. Trying Out, Evaluating and Revising the Second Draft of the First Assessment Topic.....	97
1) The Try-Out.....	97
2) The Evaluation.....	98
a) Students' Attitude toward the Second Draft Assessment Designed.....	98
b) Relation between the National Curriculum and the Second Draft Assessment Designed.....	100
c) Validity.....	100
i) Content Validity.....	101
ii) Construct Validity.....	102
iii) Face Validity.....	102
d) Reliability.....	103
e) Practicality.....	105

f) Authenticity and Interactiveness.....	106
g) Backwash Effect.....	107
h) Fairness.....	108
3) The Revision.....	109
b. Trying Out, Evaluating and Revising the Second Draft of the Second Assessment Topic.....	110
1) The Try-Out.....	110
2) The Evaluation.....	110
a) Students' Attitude toward the Second Draft Assessment Designed.....	110
b) Relation between the National Curriculum and the Second Draft Assessment Designed.....	111
c) Validity.....	112
i) Content Validity.....	112
ii) Construct Validity.....	113
iii) Face Validity.....	114
d) Reliability.....	115
e) Practicality.....	116
f) Authenticity and Interactiveness.....	117
g) Backwash Effect.....	118
h) Fairness.....	120
3) The Revision.....	120
c. Trying Out, Evaluating and Revising the Second Draft of the Third Assessment Topic.....	121
1) The Try-Out.....	121
2) The Evaluation.....	121
a) Students' Attitude toward the Second Draft Assessment Designed.....	121
b) Relation between the National Curriculum and the Second Draft Assessment Designed.....	122
c) Validity.....	123

i) Content Validity.....	123
ii) Construct Validity.....	125
iii) Face Validity.....	125
d) Reliability.....	127
e) Practicality.....	128
f) Authenticity and Interactiveness.....	129
g) Backwash Effect.....	130
h) Fairness.....	131
3) The Revision.....	132
4. The Final Draft of the Listening Assessment.....	132
a. First Assessment (<i>In the Classroom</i>).....	133
b. Second Assessment (<i>A Cup of Tea</i>).....	134
c. Third Assessment (<i>Good Daily Activities</i>).....	136
B. Discussion.....	138
CHAPTER V CONCLUSSIONS, IMPLICATIONS AND	
SUGGESTIONS.....	141
A. Conclusions.....	141
1. The Designed Listening Assessment for the Fifth Grade Students of SDN Rejodani Ngaglik Sleman Yogyakarta.....	140
2. The Characteristics of the Appropriate of the Assessment Tasks for the Fifth Grade Students of SDN Rejodani Ngaglik Sleman Yogyakarta.....	142
a. Validity.....	142
b. Reliability.....	142
c. Practicality.....	143
d. Good Backwash Effect.....	143
e. Fairness.....	143
f. Authenticity and Interactiveness.....	143
g. Relation between the Designed Assessment and the National Curriculum.....	143
B. Implications.....	144

C. Suggestions.....	145
1. The English Teachers.....	145
2. The Principals of the Elementary Schools.....	145
3. English Education Department.....	145
4. The Government.....	146
5. The Other Researchers.....	146
REFERENCES.....	147
APPENDICES.....	150

LIST OF FIGURES

	Page
Figure 1: Construct Validity of Score Interpretation proposed by Bachman and Palmer (1996: 22).....	17
Figure 2: Practicality proposed by Bachman and Palmer (1996: 36).....	19
Figure 3: Test Development.....	34
Figure 4: Conceptual Framework.....	41
Figure 5: The Mean Formula proposed by Suharto (2005: 58).....	50
Figure 6: The Range Formula proposed by Suharto (2005: 59).....	50
Figure 7: Research Procedure.....	53
Figure 8: Steps in Developing the Appropriate Listening Assessment.....	54
Figure 9 : <i>In the Classroom</i> Assessment Design.....	72
Figure 10: <i>A Cup of Tea</i> Assessment Design.....	75
Figure 11: <i>Good Daily Activities</i> Assessment Design.....	78

LIST OF TABLES

	Page
Table 1: Characteristics of English-language Learners.....	10
Table 2: The Test Structure of Cambridge Flyers 6 (2009: 4).....	22
Table 3: The Test Task Specifications of Cambridge Flyers 6 Listening Test (2009: 4).....	23
Table 4: The Test Task Specifications of Cambridge Flyers 6 Reading and Writing Tests (2009: 5).....	24
Table 5: The Test Task Specifications of Cambridge Flyers 6 Speaking Test (2009: 5).....	25
Table 6: The Areas of Language Knowledge proposed by Bachman and Palmer (1996: 68).....	32
Table 7: Observable Performance of the Four Skills proposed by Brown (2004: 118).....	37
Table 8: Blueprint of the Need Analysis Questionnaire.....	43
Table 9: Blueprint of the Evaluation Questionnaire administered to the Experts.....	44
Table 10: Blueprint of the Evaluation Questionnaire administered to the Students.....	45
Table 11: Blueprint of the Classroom Observation Sheet.....	46
Table 12: Quantitative Data Conversion.....	50
Table 13: Blueprint of the Need Analysis Questionnaire.....	55
Table 14: Data of the Fifth Grade Students of SDN Rejodani.....	56
Table 15: Students' Responses on Needs Analysis Questionnaires of the First Element by Question Number, Learners' Percentage and Category.....	56
Table 16: Students' Responses on Needs Analysis Questionnaires of the Second Element by Question Number, Learners' Percentage and Category.....	57
Table 17: Students' Responses on Needs Analysis Questionnaires of the Third Element by Question Number, Learners' Percentage and	

	Category.....	58
Table 18:	Students' Responses on Needs Analysis Questionnaires of the Fourth Element by Question Number, Learners' Percentage and Category.....	59
Table 19:	Students' Responses on Needs Analysis Questionnaires of the Fifth Element by Question Number, Learners' Percentage and Category.....	59
Table 20:	The Assessment Structure of the First Assessment.....	61
Table 21:	The Assessment Structure of the Second Assessment.....	62
Table 22:	The Assessment Structure of the Third Assessment.....	63
Table 23:	The Standard of Graduate Competence, Standard of Competence and Basic Competence of the Fifth Grade Elementary School Students and the Language Skill in National Curriculum (2008: 11, 17-19).....	64
Table 24:	The Assessment Task Characteristics of the First Assessment.....	65
Table 25:	The Assessment Task Characteristics of the Second Assessment.....	67
Table 26:	The Assessment Task Characteristics of the Third Assessment....	68
Table 27:	The Criteria of Marking of the First Assessment.....	69
Table 28:	The Criteria of Marking of the Second Assessment.....	70
Table 29:	The Criteria of Marking of the Third Assessment.....	71
Table 30:	The Assessment Specifications of the First Assessment.....	73
Table 31:	The Assessment Specifications of the Second Assessment.....	76
Table 32:	The Assessment Specifications of the Third Assessment.....	78
Table 33:	Blueprint of the Evaluation Questionnaire administered to Respondents.....	80
Table 34:	Experts' Responses on Evaluation Questionnaires of the First Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	82
Table 35:	Experts' Responses on Evaluation Questionnaires of the Second Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	83

Table 36:	Experts' Responses on Evaluation Questionnaires of the Third Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	84
Table 37:	Experts' Responses on Evaluation Questionnaires of the Fourth Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	84
Table 38:	Experts' Responses on Evaluation Questionnaires of the Fifth Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	85
Table 39:	Experts' Responses on Evaluation Questionnaires of the Sixth Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	85
Table 40:	Experts' Responses on Evaluation Questionnaires of the Seventh Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	85
Table 41:	Experts' Responses on Evaluation Questionnaires of the First Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	86
Table 42:	Experts' Responses on Evaluation Questionnaires of the Second Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	87
Table 43:	Experts' Responses on Evaluation Questionnaires of the Third Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	88
Table 44:	Experts' Responses on Evaluation Questionnaires of the Fourth Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	88
Table 45:	Experts' Responses on Evaluation Questionnaires of the Fifth Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	89

Table 46:	Experts' Responses on Evaluation Questionnaires of the Sixth Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	89
Table 47:	Experts' Responses on Evaluation Questionnaires of the Seventh Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and Category.....	90
Table 48:	Experts' Responses on Evaluation Questionnaires of the First Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	90
Table 49:	Experts' Responses on Evaluation Questionnaires of the Second Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	91
Table 50:	Experts' Responses on Evaluation Questionnaires of the Third Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	92
Table 51:	Experts' Responses on Evaluation Questionnaires of the Fourth Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	92
Table 52:	Experts' Responses on Evaluation Questionnaires of the Fifth Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	93
Table 53:	Experts' Responses on Evaluation Questionnaires of the Sixth Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	93
Table 54:	Experts' Responses on Evaluation Questionnaires of the Seventh Element " <i>Good Daily Activities</i> " by Question Number, Mean Scores and Category.....	94
Table 55:	The General Revision Table of the First Draft of the Designed Listening Assessment.....	94
Table 56:	Blueprint of the Evaluation Questionnaire administered to the	96

Table 57:	Students.....	96
Table 58:	Blueprint of the Classroom Observation Sheet.....	
	Students' Responses on Evaluation Questionnaires of the First Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	98
Table 59:	The Teacher's Responses on Observation Sheet of the First Element " <i>In the Classroom</i> " by Question Number and Statements.....	99
Table 60:	Respondents' Responses on Observation Sheet of the Second Element " <i>In the Classroom</i> " by Question Number and Statements.....	100
Table 61:	The Teacher's Responses on Observation Sheet of the Third Element " <i>In the Classroom</i> " by Question Number and Statements.....	101
Table 62:	Students' Responses on Evaluation Questionnaires of the Third Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	101
Table 63:	The Observer's Response on Observation Sheet of the Third Element " <i>In the Classroom</i> " by Question Number and Statements.....	102
Table 64:	Students' Response on Evaluation Questionnaires of the Third Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	103
Table 65:	Respondents' Responses on Observation Sheet of the Third Element " <i>In the Classroom</i> " by Question Number and Statements.....	103
Table 66:	Students' Response on Evaluation Questionnaires of the Fourth Element " <i>In the Classroom</i> " by Question Number, Mean Scores and Category.....	104
Table 67:	The Observer's Response on Observation Sheet of the Fourth Element " <i>In the Classroom</i> " by Question Number and	104

Table 68:	Statements.....	
	The Teacher's Responses on Observation Sheet of the Fourth Element " <i>In the Classroom</i> " by Question Number and	105
Table 69:	Statements.....	
	Students' Responses on Evaluation Questionnaires of the Fifth Element " <i>In the Classroom</i> " by Question Number, Mean Scores	106
Table 70:	and Category.....	
	Students' Response on Evaluation Questionnaires of the Sixth Element " <i>In the Classroom</i> " by Question Number, Mean Scores	107
Table 71:	and Category.....	
	The Teacher's Responses on Observation Sheet of the Sixth Element " <i>In the Classroom</i> " by Question Number and	107
Table 72:	Statements.....	
	The Observer's Responses on Observation Sheet of the Sixth Element " <i>In the Classroom</i> " by Question Number and	107
Table 73:	Statements.....	
	The Observer's Responses on Observation Sheet of the Seventh Element " <i>In the Classroom</i> " by Question Number and	108
Table 74:	Statements.....	
	Students' Responses on Evaluation Questionnaires of the Seventh Element " <i>In the Classroom</i> " by Question Number,	108
Table 75:	Mean Scores and Category.....	
	Respondents' Responses on Observation Sheet of the Eighth Element " <i>In the Classroom</i> " by Question Number and	109
Table 76:	Statements.....	
	The Revision of the Assessment Aspects of the First Assessment	109
Table 77:	Topic.....	
	Students' Responses on Evaluation Questionnaires of the First Element " <i>A Cup of Tea</i> " by Question Number, Mean Scores and	110
Table 78:	Category.....	
	The Observer's Responses on Observation Sheet of the Second	

	Element “A Cup of Tea” by Question Number and	111
Table 79:	Statements.....	
	The Observer’s Responses on Observation Sheet of the Third	113
Table 80:	Element “A Cup of Tea” by Question Number and Statements.....	
	Students’ Responses on Evaluation Questionnaires of the Third	
	Element “A Cup of Tea” by Question Number,Mean Scores and	113
Table 81:	Category.....	
	Respondents’ Responses on Observation Sheet of the Third	114
Table 82:	Element “A Cup of Tea” by Question Number and Statements.....	
	Students’ Responses on Evaluation Questionnaires of the Third	
	Element “A Cup of Tea” by Question Number,Mean Scores and	114
Table 83:	Category.....	
	The Teacher’s Responses on Observation Sheet of the Third	115
Table 84:	Element “A Cup of Tea” by Question Number and Statements.....	
	The Teacher’s Responses on Observation Sheet of the Fourth	115
Table 85:	Element “A Cup of Tea” by Question Number and Statements.....	
	Students’ Responses on Evaluation Questionnaires of the Fourth	
	Element “A Cup of Tea” by Question Number,Mean Scores and	116
Table 86:	Category.....	
	Students’ Responses on Evaluation Questionnaires of the Fifth	
	Element “A Cup of Tea” by Question Number,Mean Scores and	117
Table 87:	Category.....	
	The Observer’s Responses on Observation Sheet of the Fifth	117
Table 88:	Element “A Cup of Tea” by Question Number and Statements.....	
	Students’ Responses on Evaluation Questionnaires of the Sixth	
	Element “A Cup of Tea” by Question Number,Mean Scores and	118
Table 89:	Category.....	
	Students’ Responses on Evaluation Questionnaires of the	
	Seventh Element “A Cup of Tea” by Question Number,Mean	119
Table 90:	Scores and Category.....	
	The Teacher’s Responses on Observation Sheet of the Seventh	119

Table 91:	Element “ <i>A Cup of Tea</i> ” by Question Number and Statements.....	
	The Teacher’s Responses on Observation Sheet of the Eighth	120
Table 92:	Element “ <i>A Cup of Tea</i> ” by Question Number and Statements.....	
	The Revision of the Assessment Aspects of the Second	121
Table 93:	Assessment Topic.....	
	Students’ Responses on Evaluation Questionnaires of the First	
	Element “ <i>Good Daily Activities</i> ” by Question Number,Mean	121
Table 94:	Scores and Category.....	
	Respondents’ Responses on Observation Sheet of the Second	
	Element “ <i>Good Daily Activities</i> ” by Question Number and	123
Table 95:	Statements.....	
	The Teacher’s Responses on Observation Sheet of the Third	
	Element“ <i>Good Daily Activities</i> ” by Question Number and	124
Table 96:	Statements.....	
	The Observer’s Responses on Observation Sheet of the Third	
	Element“ <i>Good Daily Activities</i> ” by Question Number and	124
Table 97:	Statements.....	
	Students’ Response on Evaluation Questionnaires of the Third	
	Element “ <i>Good Daily Activities</i> ” by Question Number, Mean	125
Table 98:	Scores and Category.....	
	The Teacher’s Responses on Observation Sheet of the Third	
	Element“ <i>Good Daily Activities</i> ”by Question Number and	125
Table 99:	Statements.....	
	Respondents’ Responses on Observation Sheet of the Third	
	Element “ <i>Good Daily Activities</i> ” by Question Number and	126
Table 100:	Statements.....	
	Students’ Response on Evaluation Questionnaires of the Third	
	Element “ <i>Good Daily Activities</i> ” by Question Number, Mean	126
Table 101:	Scores and Category.....	
	The Observer’s Responses on Observation Sheet of the Fourth	
	Element “ <i>Good Daily Activities</i> ” by Question Number and	127

Table 102:	Statements.....	
	Respondents' Responses on Observation Sheet of the Fifth Element " <i>Good Daily Activities</i> " by Question Number and	128
Table 103:	Statements.....	
	Students' Response on Evaluation Questionnaires of the Fifth Element " <i>Good Daily Activities</i> " by Question Number, Mean	128
Table 104:	Scores and Category.....	
	Respondents' Responses on Observation Sheet of the Sixth Element " <i>Good Daily Activities</i> " by Question Number and	129
Table 105:	Statements.....	
	Students' Responses on Evaluation Questionnaires of the Sixth Element " <i>Good Daily Activities</i> " by Question Number, Mean	130
Table 106:	Scores and Category.....	
	Students' Response on Evaluation Questionnaires of the Seventh Element " <i>Good Daily Activities</i> " by Question Number, Mean	130
Table 107:	Scores and Category.....	
	Respondents' Responses on Observation Sheet of the Eighth Element " <i>Good Daily Activities</i> " by Question Number and	131
Table 108:	Statements.....	
	The Revision of the Assessment Aspects of the Third	132
Table 109:	Assessment Topic.....	133
Table 110:	The Final Draft of the First Assessment Specifications.....	134
Table 111:	The Final Draft of the Second Assessment Specifications.....	135
	The Final Draft of the Third Assessment Specifications.....	

LIST OF APPENDICES

	Page
Appendix A: Data of Students and Respondents.....	151
Appendix B: Blueprint of the Designed Listening Assessment.....	153
Appendix C: Research Instruments.....	161
Appendix D: Quantitative Data Analysis.....	183
Appendix E: Qualitative Data.....	201
Appendix F: The First Draft.....	229
Appendix G: The Second Draft.....	253
Appendix H: The Final Draft.....	279

DESIGNING ENGLISH LISTENING ASSESSMENT FOR THE FIFTH GRADE STUDENTS OF SDN REJODANI NGAGLIK SLEMAN YOGYAKARTA

By:
Amanda Widyati
NIM 07202241026

ABSTRACT

The objective of the research was to design an appropriate listening assessment for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta. The products were a set of listening assessment worksheets, the audio recording and the teacher's guide.

The research study was categorized as Educational Research and Development (R & D). The steps were conducting need analysis, writing the blueprint of the listening assessment, designing the first draft of the listening assessment, getting expert judgement, revising the first draft, designing the second draft, trying-out the second draft, revising the second draft, and writing the final draft of the listening assessment. The research instruments were questionnaires, observation sheet, interview guide and the assessment marks. The needs analysis and evaluation questionnaire had fulfilled the content validity and the reliability was measured through the Cronbach Alpha's statistics. The observation sheet had fulfilled the content validity and the external audit. The interviews accomplished the content validity and the external audit in coding the responses. The questionnaires were used to get the data of students' needs and the data to evaluate the designed listening assessment from the students and the experts. The observation sheet was used to get the data from the English teacher and the observer during the first try-out. The interview sessions were done with the one of the students and the teacher after the first try-out. The reliability of the assessment was measured through the test-retest method.

The result of the research study shows that: (1) the designed listening assessment for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta is colourful and has a lot of pictures and (2) the characteristics of the appropriate listening assessment for the fifth grade students of SDN Rejodani are valid, reliable, practical, have good backwash effect, fair, authentic and interactive and made based on the students' needs and the National Curriculum.

CHAPTER I

INTRODUCTION

A. Background of the Problem

Teaching English to young learners is not a new issue in the world of education. Some of children's characteristics are considered having good contribution for them in learning English. They are curious to know everything and comfortable with routines. Moreover, they are also sensitive to the new language sounds and rhythm that can support their language learning(Pinter, 2006).

The process of teaching and learning English to children cannot be separated from the assessment. The assessment itself is described as an effort to analyze the students' progress and achievement on teaching and learning process. As the integral part of the teaching and learning, assessment is important for some reasons. First, the assessment can evaluate the students' knowledge and students' needs. Second, because some students might learn faster or slower than the others, the teacher can check their progress through the regular assessment. Third, the assessment can motivate students. The result of the assessment may help the students to understand how capable they are in a lesson. In case of the good result, it will motivate them to maintain their achievement while in case of the bad result; assessment can motivate the students to achieve a better result by learning more. The last, assessment functions as a media for teacher's self-evaluation. The teacher can help students more to master the lesson by exploring the new

techniques and trying to find out the materials that appropriate for the students' needs.

However, sometimes the assessment cannot measure the students' competence successfully. The factors can come from the inside and outside the test-takers. Some of the students may fail although they have capability in doing the test. In this case, the assessor should consider the inside factors from the students like their health, psychology, etc. The causes also may come from the outside the students like the type of the assessment, the time, the test environment, etc.

Based on the observation, the assessment of the four language skills is done in some ways in this school. First, LKS is as the only source to assess reading skill. Reading a short text, finding the difficult words and answering the comprehension questions of the text are the activities that should be done by the students. Second, dictation is used as the way to assess writing skill. Students are asked to write down the words with correct spelling. Third, speaking assessment in this school is never done individually. Imitating teacher's words for the whole class is the activity and the score is given based on the students' pronunciation. The last is listening assessment which is done by translating words.

This research study is concerned with the listening assessment since it is as the first important part before the students learn to read, write, and speak. Listening as the receptive skill enables students to get a lot of input and the new knowledge of language. Some of the listening sub-skills like inferring meaning from context and listening for specific information can increase the students' ability in other skills (Georgiou & Pavlou, 2003). It may help the students to get the earlier start in their

language learning process. In this case, listening assessment is used to measure the students' achievement for the input of language that's given. The result can be used as a reference to develop the students' abilities in other skills.

As what have been mentioned above, the listening assessment has already existed in this school in the form of translation test. However, the content, the form and the methods are inappropriate. Therefore, it is necessary to conduct a research on designing an appropriate English listening assessment in SDN Rejondani Sleman Yogyakarta.

B. Identification of the Problem

In the background of the problem, it is clearly stated that listening assessment is needed to monitor and evaluate students' achievement which can support the students' abilities in other language skills. Those functions can be optimally achieved when the principles of a good assessment are fulfilled. They are validity, reliability, practicality, authenticity and interactiveness, having good backwash effect and fairness. Those aspects have close relationship with the teacher and students as the important elements of the teaching and learning process. Therefore, there are some factors which give contribution to the listening assessment process in this school.

A good listening assessment should be valid and reliable. Those criteria mean that the assessment assesses what should be assessed and has the scoring guidance to avoid bias. In case of the listening assessment in this school, those criteria cannot be investigated because the form of the listening assessment is inappropriate and there is no rubric of the assessment.

Practical means that the listening assessment should be feasible to do. Time, money, space and assessors are some of the aspects that need to be considered. In this school, the students translate some words as the listening assessment because the listening assessment's method which can accommodate all of the students in the classrooms is not applicable. The teacher's responsibility for handling all of the English classes limits the teacher's time to design the assessment for her students.

Authenticity and interactiveness in a good listening assessment are related to the appropriateness of the assessment materials for the students. Materials in good listening assessment should have related with the students' daily life. In this school, the listening assessment does not fulfil the criteria of authentic and interactive. The reason is because the listening assessment which is done by the teacher cannot be found in the real life of the students.

A good listening assessment should give motivation to the students. Whether the result is good or bad, students can learn and get the feedback to improve their listening skill in the future. This called as good backwash effect. What happened in this school is that the students are afraid to face the next assessment after they had the previous one because it only tests their vocabularies. The teacher dictates some words in English then ask the students to rewrite the words in Indonesian. The impact is that the students are worried that they do not have enough vocabularies which can make them failed in the test. The time which given also makes the students are afraid. For each word, they were given only about two

minutes to think the answer. This create the unfriendly atmosphere because very limited time is really intimidating the students when doing the test.

Fairness is the last principle of a good assessment. All of the students should get the same chance to do the assessment as good as they can. In this school, the example of the unfairness is the students at the back cannot hear the teacher's voice when she dictates the words to translate. Sometimes, they submit the answer sheet without translating all of the words. It doesn't mean they do not know the answers but it's because of the unfairness that the students at the back cannot hear the teacher's voice. They do not have the same chance as the other students in the classroom.

There are also some problems related to the students and teacher in this school. A lot of students in each class make the teacher are difficult to manage her class. Moreover, as what have been mentioned above, the teacher has to handle all of the English class in this school. This condition will make her exhausted so it's possible that she cannot teach and assess her students optimally.

C. Limitation of the Problem

Based on the identification of the problem, the researcher limits the research study on the technical quality of the good listening assessment because of two reasons.

First, the researcher has limited time in doing the research. Second, the researcher doesn't have access to obtain all the feasible facilities that are required to solve all of the problems. The last, the problem of listening assessment in SDN

Rejodani has represented the problems in other elementary schools in the same area.

D. Formulation of the Problem

This research study is concerned with the listening assessments for the fifth grade students of SDN Rejodani. The problems are formulated as follows:

- a. What do the designed listening assessment for the fifth grade students of SDN Rejodani look like?
- b. What are the characteristics the appropriate listening assessment for the fifth grade students of SDN Rejodani?

E. Objective of the Study

Based on the problem formulation above, the purpose of the study is to design an appropriate listening assessment for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta. At the end of the study, the researcher is expected to produce a set of listening assessment.

F. Significance of the Study

This study is expected giving some advantages in some ways. First, the result of the study can give the new knowledge to the English teacher in SDN Rejodani Sleman Yogyakarta to assess her students appropriately. It can also encourage other researchers to conduct similar research with different skills and group of students.

CHAPTER II LITERATURE REVIEW

A. Theoretical Review

1. Teaching English to Young Learners

Introducing foreign language learning to young language learners will give some benefits for them. First, it can be used as the basic when they continue to the secondary school. Second, some of young learners' characteristics may support their language learning in the earlier stages (Brewster & Ellis, 2002). One of them is that they are active learners. They also have positive attitude about learning and able to work with others (Scott & Ytreberg, 1990).

Considering some of the facts above, Indonesian government has introduced English in Elementary School as the local content subject since 1994. It is still as the local content subject till now as what have been mentioned in 2004 National Curriculum (Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar, 2008). To make the process of teaching and learning English in elementary school is suitable for the learners' need and their development, the government launched *Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar* as guidance for all of the elementary school teachers in Indonesia.

2. Assessment for Young Language Learners

Assessment and the process of teaching and learning are equally important. As the integrated part, both of them give a lot of contribution to the learners' achievement. Linse (2005: 118) proposed some of principles to assess young language learners as what have been listed below:

1. Assessment should be seen from a learning-centred perspective.
2. Assessment should support learning and teaching.

3. Assessment is more than testing.
4. Assessment should be congruent with learning.
5. Children and parent should understand assessment issues.

This part has discusses about the assessment, its types and the principles. Cambridge Test for Young Learners as the standardized test for young learners and young language learners as the test-takers are discussed together with the steps in designing the assessment.

a. The Nature of Assessment

Assessment is the process of collecting information related to the learners' achievement. According to Brewster and Ellis (2002), assessment can be described as an attempt from the assessors to analyze the learning that the learners have achieved for a period of time as a result of the teaching and learning process. In another book, Pinter (2006) states the similar definition about assessment which refers to the data analysis that used by the teacher as the evidence of the learners' performance and progress. Therefore, it is clear that assessment is used as a media to evaluate the learners' achievement. However, the successful of any assessment depends on the selection and the appropriateness of the assessment tools and procedures as well as the proper interpretation of the learners' performance(Shaaban, 2001).

The importance of assessment can be seen through some aspects. First, assessment can be used as guidance for the students in their study (Wakeford, 2003). Students can use the result of the assessment to improve their competencies in the future. Second, the effectiveness of teaching methods can be monitored through the assessment. It is important to plan the teacher's future work. The information may help the teacher to modify the aspects of the teaching

and develop the techniques for responding to the young learners' individual needs (Georgiou & Pavlou, 2003). The last, assessment can provide the information about the learners to their parents. They need to know whether their child's progress achieve the satisfactory results or not. The reason is that they want to be involved in and support their child's language learning (Brewster & Ellis, 2002).

b. Young Learners as the Test-Takers

Assessment is not only needed by the advanced learners but also the beginners in primary schools. However, the assessors should pay attention to their different needs and characteristics to decide the suitable techniques in assessing them.

1) The Characteristics of Young Language Learners

Young learners bring their own personalities, individual cognitive styles and capabilities, and also their strengths and weaknesses in their language learning (McKay, 2006). They have their own uniqueness that helps them to develop their competencies.

Young language learners can be divided into two groups based on their range of ages. They are very young learners (VYLs) and young learners (YLS) (Slattery & Willis, 2001). Children who belong to very young learners are the children under seven years old. Meanwhile, young learners are the children who are seven to twelve years old.

Very young learners are easy to pick up the new language in everyday situations in their environment (Pinter, 2006). They can learn better when they are involved in the activities which provide songs, games and stories because they love to learn in fun ways. Moreover, Brewster and Ellis (2002: 27) state "children have a lot of physical energy and often need to be physically active". That is why

preparing a lot of activities are important. Another very young learners' characteristic is that they have short attention span (Scott & Ytreberg, 2004). Children are easy to get bored. Doing the various activities can make them still happy and active in the classroom. However, young learners have quite different characteristics. Slattery and Willis (2001: 5) describe some characteristics of children between seven and twelve years old (Ys) as follows:

1. learn to read and write in their own language
2. develop as thinkers
3. understand the difference between the real and the imaginary
4. can plan and organize how best to carry out the activity
5. can work with others and learn from others
6. can be reliable and take responsibility for class activities and routines.

Both VYs and Ys may come from different literacy backgrounds. It should be considered to develop the appropriate assessment. Below are the categories of English language learners adapted from Freeman and Freeman (2003).

Table 1: Characteristics of English-language Learners

Newly arrived students with adequate formal schooling	Newly arrived students with limited formal schooling	Students exposed to two languages simultaneously	Long-term English-language learners
<ul style="list-style-type: none"> - Have been in the country for fewer than five years, - Have difficulty with standardized tests, - Have parents who are educated speakers of the L1, - Developed a strong foundation in their L1, - Demonstrate the potential to make fast progress in English, - Have found it easy to acquire a second or third language. 	<ul style="list-style-type: none"> - Have recently arrived in an English-speaking school (fewer than five years), - Have limited native language and literacy skills, - Perform poorly on achievement tasks, - Could have difficulty learning English. 	<ul style="list-style-type: none"> - Have grown up being exposed to two languages simultaneously, - May have not developed academic literacy in either L1 or L2 - Live in communities of speakers who primarily communicate in their L1 or go back and forth between languages. 	<ul style="list-style-type: none"> - Have already spent more than five years in an English-speaking school, - Have had some English as a second language classes or bilingual support, - Require substantial and ongoing language and literacy support.

2) Issues on Assessing Young Learners

As what have been mentioned above, young learners are different from adult learners. The difference is not in the teaching and learning process only but also in terms of the assessment. To provide the suitable assessment for young learners, it requires the commitments from the policymakers and practitioners to improve the existing assessment by develop and implement appropriate assessment tools and procedures (Garcia, Lawton, & Diniz de Figueiredo, 2010). Here are some issues related to the assessment of young learners.

One of the issues is the traditional methods of assessment to young learners. The traditional methods of assessment which usually rely on ‘pencil and paper test’ are considered inappropriate for young learners. The method cannot show what the learners know and what the learners can do since they are often not very good yet in writing and it can make them stressful and tiring (Pinter, 2006). The examples of the traditional method assessment are answering multiple choice questions and filling in the gaps in sentences.

The assessment environment is also important for young learners. Assessing them in friendly situation will increase their motivation in learning English. Georgiou and Pavlou (2003) promote some of the assessment techniques like structured assessment tasks, portfolio assessment and projects which are child-friendly that can keep the learners’ positive attitude toward English.

Linse (2005, 141-142) also promotes some aspects except validity and reliability that should take into account before selecting the appropriate tools for assessing young learners. They are reasonable expectations, wait time, transitioning into English and instructions to the learners.

1. “Reasonable Expectations”

The expectations of the young learners should make sense. When the children still can only use some words or short sentences to express something in English, it will be better to avoid such unrealistic expectations like they can use English in longer sentences or even fluent in English.

2. “Wait Time”

Assessing young learners needs patience. The teacher should give them enough time to process the information. Interrupting them only break their concentration and can make the assessor getting the wrong conclusion.

3. “Transitioning into English”

Before starting the assessment, it will be better for the teacher to give some activities in English first like singing songs or tell riddles in English. The aim is to make sure that the learners are ready with the environment because the young learners need time to have some transitions form their native language to English.

4. “Instructions to Learners”

The teacher should make sure that the assessment instructions are clear enough for the learners. The instructions have to be simple and provide the complete information for them.

c. **Types of Assessment**

As the instrument to measure the learners’ competencies, assessment can be divided into three types. It can be described based on the techniques, the time implementation and the reference.

1) Assessment based on the Techniques

There are two kinds of the assessment according to the techniques. They are formal assessment and informal assessment.

a) Formal Assessment

Formal assessment can be described as the tests given under conditions that ensure the assessment of individual performance in any given area (Richard & Schimdt, 2002). In his book, Brown (2004: 6) states ‘formal assessments are exercises or procedures specifically designed to tap into a storehouse of skills and knowledge.’ When the learners have already known that the task they are doing is for assessment purposes, it is called as the formal assessment. The examples are final exam, mid-term exam, etc.

There are also some arguments which against to the formal assessment as have been listed by Mc Cann and Harris (1994: 26).

1. *‘Testing takes up time which I could better use for teaching.’*
2. *‘I know how my students are doing anyway. Why do I need to test them as well?’*
3. *‘Some of students in my class, who have worked hard all year, fail tests. Others who have done less work past them, just because they have been to an English speaking country or because they have private classess.’*
4. *‘Tests do not motivate my students at all. The good ones feel complacent and the bad ones just give up.’*

b) Informal Assessment

Informal assessment can be done in some ways. Complementing the learners like ‘Good job!’ can be categorized as the informal assessment. Informal assessment can also be done by collecting the data about the students’ performance is done under normal classroom conditions instead of the using of standardized tests (Richard& Schimdt, 2002). Therefore, the form of informal assessment comes from the unplanned comments and responses from the teacher.

2) Assessment based on the Time Implementation

In this part, assessment is divided into two types considering when it will be implemented. They are formative and summative assessment.

a) Formative Assessment

There are some definitions about formative assessment according to some experts. Cameron (2001) states the purpose of formative assessment is to give the data of on-going teaching and learning accompanying by the immediate feedback. Therefore, formative assessment is used as a media to evaluate the learners' achievement which provides information for the teacher's future plan of materials, media or methods. Formative assessment is also defined as the learners' evaluating process of competencies and skills with the goal of helping them to continue their forming process of their competencies and skills (Brown, 2004).

b) Summative Assessment

Summative assessment is quite different from the formative assessment. It aims to measure or summarize the learners' competence at the end of a course or unit of instruction (Brown, 2004). The assessment is administered at the end of the course and the teachers can see whether the objectives which set out have achieved or not. Harris and Mc Cann (1994) said that obtaining a lot of information from the summative assessment does not give feedback into the learning process. In her book, Cameron (2001) has the same opinion that the summative assessment does not give impact on the next round of teaching.

3) Assessment based on the References

There are two kinds of assessment based on the references. They are norm-referenced and criterion referenced assessment.

a) Norm-Referenced Assessment

In norm-referenced assessment, Brown (2004: 7) states “each test-taker’s score is interpreted in relation to mean (average score), median (middle score), standard deviation (extent of variation in scores), and/or percentile rank”. When the teacher assesses the learners’ achievement and he/she create a ranking which can give information whether the learners learn more or less than the others or not, it is called as norm-referenced assessment (Cameron, 2001). This kind of assessment may motivate some learners to do better than their peers and move up the ranking. The result of the assessment is interpreted by comparing one of the learners with the others based on the ranking that have been made.

b) Criterion-Referenced Assessment

Criterion-referenced assessment can match the child’s performance against an expected response on an item (Cameron, 2001). In this kind of assessment, the learners’ performance relies on the expected norms. It is designed to give the test-takers the feedback in the form of grades on specific course or lesson objectives (Brown, 2004).

d. Principles of Language Assessment

Good language assessment must fulfil some criteria. The aspects which influence the achievement of the criteria can come from inside and outside the assessment itself. Here are the criteria that should be achieved by a good language assessment.

1) Validity

Assessment should be valid. It means that when designing an assessment, the assessment developers need to think about what they want to assess, what the

proposed assessment will assess, and what can be claimed from the outcomes of the assessment. At least, a good assessment needs to fulfil the three types of validity. They are content validity, construct validity and face validity.

a) Content Validity

Content validity can be indicated from the aspect of whether the assessment assesses what supposed to be assessed or not. If the assessment has already assessed what it claims to, usually the achievement that is measured can be clearly defined. In this case, content validity can be achieved not only in the process of making the assessment but also in the process of interpreting the result.

Content validity also can be seen through how well the assessment includes all of the components that want to be assessed and whether it includes some of other components outside or not. According to Brown (2004), people can measure content validity through identifying and framing the classroom objectives and checking whether the lesson objectives have already represented in the form of test specifications or not.

b) Construct Validity

Construct validity refers to the possibility of an assessment procedure that is intended to produce a measure of a particular construct (Children, C. o., 2008).

Sometimes, it is very difficult to assess one of the competencies without also assessing others. In his book, Baxter (1997) states the question whether the assessment really assesses what it's supposed to assess and nothing else to check construct validity.

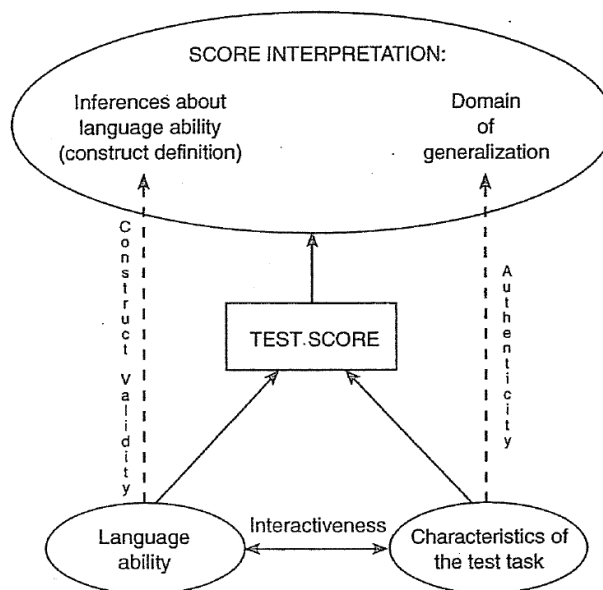


Figure 1: Construct Validity of Score Interpretation proposed by Bachman and Palmer (1996: 22)

It is also used to refer to the extent of which the people can interpret a given assessment score as an indicator of the abilities or constructs that want to be measured. According to Bachman and Palmer (1996), construct validity can be defined as the meaningfulness and the appropriateness of the interpretation that was made for the basis of test scores.

Construct validity does not focus on the assessment procedure and the score interpretation only. It also concerns on the nature of the broader constructs tested like recall of knowledge and the demonstration of teamwork skills (Wakeford, 2003).

c) Face Validity

Face validity is something to do with the layout of the assessment. As the example, people can check the face validity through the pictures in the assessment which are clear enough or not for the test-takers. Brown (2004: 33-34) also

describes some aspects that will be generally considered fulfil the face validity by the test-takers as follows:

1. The directions are clear,
2. the structure of the test is organized logically,
3. its difficulty level is appropriately pitched,
4. the test has no “surprise”, and
5. the timing is appropriate.

Face validity also means that the assessment appears to test what is trying to test (Baxter, 1997). It is almost always perceived in terms of the content. If the test samples the actual content of what the test-takers have achieved or experts to achieve, then the face validity will be perceived.

2) Reliability

Another important aspect of the assessment is reliability. Reliability measures how well a test or assessment assesses what it claims. According to Bachman and Palmer (1996), it is often described as consistency of measurement. A reliable assessment provides consistency of the data when the assessment is given to different test-takers with the similar background and the result of the assessment will give the similar scores too when it is given to different assessors with the similar background also. In his book, Baxter (1997) defined the two aspects above as the test reliability and the scorer reliability. To increase the reliability of the assessment, it will be very useful if the assessment provides the guidelines for scoring. By providing scoring rubrics and unambiguous scoring guidelines, it can make the assessor is easier to interpret the test results accurately and to assign correct scores to the test-takers (Gordon, 2007).

According to Brown (2004: 31), the assessment reliability depends on the physical context of the assessment itself. As the assessor, some aspects below can enhance the certainty that the assessment is reliable.

1. Every student has a cleanly photocopied test sheet,
2. Sound amplification is clearly audible to everyone in the room,
3. Video input is equally visible to all,
4. Lighting, temperature, extraneous noise, and other classroom conditions are equal (and optimal) for all students, and
5. Objective scoring procedures leave little debate about correctness of an answer.

3) Practicality

A well-made assessment should be feasible to do. Some of the elements that should be considered are the time which is given to the test-takers, the materials and equipments, the cost of the assessment, and whether the scoring system is feasible in the teacher's time frame or not.

According to Bachman and Palmer (1996), to determine the practicality of the assessment, there are two important aspects. First, the resources that will be needed to develop an assessment and the second is the allocation and management of the resources that are available. The relationship can be presented as follows:

$$Practicality = \frac{Available\ resources}{Required\ resources}$$

If practicality ≥ 1 , the test development and use is practical.
If practicality < 1 , the test development and use is not practical.

Figure 2: Practicality proposed by Bachman and Palmer (1996: 36)

4) Authenticity and Interactiveness

Authenticity is something to do with the relationship between the assessment task and the Target Language Use (TLU). Bachman and Palmer (1996) defined

authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a TLU task. Therefore, a language assessment should have correlation with the situation that is faced by the test-takers. Brown (2004: 28) promotes some characteristics of the authentic assessment as follows:

1. The language in the test is as natural as possible.
2. Items are contextualized rather than isolated.
3. Topics are meaningful (relevant, interesting) for the learner.
4. Some thematic organization to items is provided, such as through a story line or episode.
5. Tasks represent, or closely approximate, real-world tasks.

Interactiveness is the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task (Bachman & Palmer, 1996). In this case, the assessment developer should be able to link up the test takers' language development with the assessment task. It can be achieved by building the interaction between the test takers as the language user and the assessment task in the TLU situation (Buck, 2001).

5) Backwash Effect

Any kind of assessment must have impact whether it is good or bad. The kind of the impact depends on the way the assessment is done. Backwash effect of the assessment is not only affected the individuals but also the educational system as well (Bachman and Palmer, 1996). One of the ways to enhance the positive backwash effect is by concerning on what happens before and after the assessment is implemented. Giving enough time to the test-takers before implementing the assessment will give them the time to review and focus to the objectives of the assessment test. Meanwhile, discussing the result of the assessment and discovering the learners' strengths and weaknesses also can enhance the beneficial

backwash effect. According to Cameron (2001), assessment as the part of the young learners' learning experience may influence the young learners whether they continue their foreign language learning or not and whether they lose their interest and motivation to learn or not.

6) Fairness

Assessment should be valid, reliable and practical but it is not enough. The aspect of fairness in an assessment is also important. The principles of fairness require the test developers to develop the assessment which give plenty of chances for the young learners to do and be able to assess them in multiple methods (Cameron, 2001). In terms of scoring methods and recording the assessment outcomes, it needs also to be fair. Wakeford (2003: 45) gives an example that 'women appear to perform less well than men in multiple choice type tests, for example, with the reverse difference for free response items, such as essays'. As the test developers, it is needed to consider a kind of that fact to avoid the unfairness of the test.

e. Cambridge Test for Young Learners as the Standardized Test for Young Language Learners

Standardized test is kind of test which presupposes certain standard objectives or criteria that held constant across one form of the test to another (Brown, 2004). The examples of standardized testing are Test of English as Foreign Language (TOEFL), International English Language Testing System (IELTS), Graduate Management Admission Test (GMAT), etc. Brown (2004) also describes some characteristics of the standardized test that can be seen as follows:

1. It has large-scale criteria which are functioned to apply a broad band of competencies that are not usually exclusive to one particular curriculum.
2. It is made through the process of empirical research and development which the goal is to place the test-takers on a continuum across a range of scores and to differentiate test-takers by their relative ranking.
3. It specifies a set of competencies for a given domain and has been constructed through a process of construct validation which will measure those competencies.

Cambridge Test for Young Learners (CTYL) is also one of the standardized tests which given to young language learners. It offers an elementary-level testing system for learners of English between the ages of seven and twelve. There are three levels of proficiency which are *Starters*, *Movers* and *Flyers*. *Starters* is usually taken by children from the age of seven. *Movers* is taken by the eight to eleven years old children and *Flyers* is taken by children who are between nine and twelve years old.

Here is the example of the test structure and the test task specifications from CTYL. It is Cambridge Flyers 6. The test is about one hour which consists of the four of language skills as what have listed in the table below:

Table 2: The Test Structure of Cambridge Flyers 6 (2009: 4)

	Length	Number of Parts	Number of Items
Listening	approx. 25 minutes	5	25
Reading and Writing	40 minutes	7	50
Speaking	approx. 7-9 minutes	4	-

The purpose of this test is to focus on ‘here and now’ context and to provide the use of language in meaningful context for children. The four language skills which tested in Cambridge Flyers 6 will be described as follows:

1) Listening

There are five parts in the listening test which consists of five items for each part. There will be an example for each part. To complete this section, the test-takers are asked to bring their own coloured pencils to mark their responses. The test task specification is listed as follows:

Table 3: The Test Task Specification of Cambridge Flyers 6 Listening Test (2009: 4)

Part	Main Skill Focus	Input	Expected Response/ Item Type	Number of Items
1	the main skill focus in all five parts of the listening test is listening for specific information of various kinds, e.g. numbers, describing people, etc.	pictures, names, dialogues	draw lines to match names to people in the picture	5
2		form or page of notepad with missing words and dialogue	write words or numbers in gaps	5
3		pictures sets and list of illustrated words or names and dialogues	match pictures with illustrated word or name by writing letter in the box	5
4		3-option multiple-choice pictures and dialogues	tick boxes next to the correct picture	5
5		picture and dialogue	carry out instructions to colour, draw and write (range of colours is: blue, brown, green, orange, pink, purple, red, yellow)	5

2) Reading and Writing

There are seven parts in reading and writing tests which have various numbers of items for each part. There will be an example for each part. The test-takers are asked to bring their pens or pencils in any colours to complete this section. The test task specification is listed as follows:

Table 4: The Test Task Specification of Cambridge Flyers 6 Reading and Writing Tests (2009: 5)

Part	Main Skill Focus	Input	Expected Response/ Item Type	Number of Items
1	reading definitions and matching to words writing words	nouns and definitions	copy the correct words next to the definitions	10
2	reading and understanding sentences	pictures and sentences	write 'yes' or 'no'	7
3	reading and completing a continuous dialogue	half a dialogue with responses in a box	select correct response and write A-H in gap	5
4	reading for specific information and gist copying words	gapped text with words in box	write words in gaps tick a box to choose the best title for the story	6
5	reading and understanding a story completing sentences	story, picture and gapped sentences	complete sentences by writing 1, 2, 3, or 4 words	7
6	reading and understanding a factual text copying words	gapped text and 3- option multiple choice (grammatical words)	complete text by selecting the correct words and copying them in corresponding gaps	10
7	reading and understanding a short text (e.g. page from diary or letter) providing words	gapped text	write words in gaps no answer options given	5

3) Speaking

In terms of the speaking test, the test takers will have to speak with one examiner about eight minutes. Before having the test, the test-takers will get a clear explanation about the format of the test by their familiar person like the teacher in their native language. This person will take the test-taker to the exam

room and introduces him/ her to the examiner. The assessment of the speaking skill is based on some criteria including comprehension, the ability to produce a prompt, appropriate and accurate response, and pronunciation. The test task specification is listed as follows:

Table 5: The Test Task Specification of Cambridge Flyers 6 Speaking Tests (2009: 5)

Part	Main Skill Focus	Input	Expected Response
1	understanding statements and responding with differences	two similar pictures oral statements about unseen picture	identify six differences in Candidate's pictures from oral statements about Examiner's picture
2	responding to questions with short answer forming questions to elicit information	one set of facts and one set of question cues	answer and ask questions about two people, objects or situations
3	understanding the beginning of a story and then continuing it based on a series of pictures	picture sequence	describe each picture in turn
4	understanding and responding to personal questions	open-ended questions about candidate	answer personal questions

f. Steps to Design an Assessment

Designing the assessment should consider the principles that have been mentioned above. According to Bachman and Palmer (1996), there are three steps of designing the test which can be adapted to develop an assessment. They are the designing stage, the operationalization stage and the administration stage.

1) The Designing Stage

The tests components should be described in detail in the designing stage. The function is to give deep information about the performance on the assessment

tasks whether it corresponds as closely as possible to the language use or not. The optimization of the test scores usefulness also can be investigated.

Bachman and Palmer (1996: 88) describe that the product of the design stage consists of the following components:

1. a description of the purpose(s) of the tests,
2. a description of the TLU domain and task types,
3. a description of the test takers for whom the test is intended,
4. a definition of the construct(s) to be measured,
5. a plan for evaluating the qualities of the usefulness, and
6. an inventory of required and available resources and a plan for their allocation and management.

Those six components above will be described briefly below:

1. The purpose of the test

The descriptions of the test purpose(s) give the information about the intension of the test. It consists of the clearly specific language ability that want to be assessed and capacity of language use that used as the basic of the test results.

2. The description of the task in TLU domain

The description of the task in TLU domain provides a set of the Target Language Use (TLU) types. The information will be as a basic to develop tasks which can be monitored the potential authenticity and the interactiveness.

3. The characteristics of the test-takers

The test-takers should be described clearly to get the assessment that is suitable for them. There are four categories which are relevant to describe the characteristics of the test-takers. They are personal characteristics, general level and profile of their language ability, the topical knowledge and the predictions of the test-takers' potential affective responses to the assessment.

4. The description of the construct(s) to be measured

The part presents the descriptions of the ability that is measured. The theoretical definition of the construct as the content of the part can be taken from the theory of language, a syllabus specification or both. Description of the construct gives clear explanations about what areas of the language abilities focus on.

5. Developing quality of the usefulness

A good assessment should fulfil some of the specific criteria like validity and reliability. In order to judge whether the assessment is good or not, making a plan to check the quality of the usefulness becomes important. Bachman and Palmer (1996) state a plan for evaluating the assessment should consist of the qualities of the usefulness and the setting of the minimum acceptable levels for each of them.

6. Identifying resources and the plan of its allocation and management

The last part has close relationship with the practicality of the assessment. Identifying resources (human, material, time) is important to check whether the assessment is feasible to do or not. The management of the resources also should be described and clearly calculated.

2) The Operationalization Stage

The operationalization stage is the second step to develop the language assessment test. In this step, the assessment test developers develop and write the actual assessment tests, writing instructions and specify the procedures for scoring the tests. The product is in the form of assessment test's blueprint which can figure out how the assessment test will be organized to form the actual tests.

There are some functions of making blueprint of the assessment proposed by Bachman and Palmer (1996: 177). They are:

1. To permit the development of other tests or parallel forms of the test with the same characteristics.
2. To evaluate the intentions of the test developers.
3. To evaluate the correspondence between the test as developed and the blueprint from which it was developed.
4. To evaluate the authenticity of the test.

The blueprint consists of two important parts. They are the test structure and test task specifications. Test structure will give information about the whole form of the test while the test task specifications describe in detail the information related to the specific purpose and the construct definition for each type of the test tasks. This is used as a template to develop the actual test tasks.

a) The Test Structure

At least there are five components in the test structure (Bachman & Palmer, 1996: 51). These five components figure out how the parts of the test are put together and presented to the test-takers.

1. “Number of Parts/ Tasks”

This part describes the number of the parts and/or tasks in a whole test.

2. “Salience of Parts/ Tasks”

In what extents all of the parts and/or tasks are clearly distinguished from one to another.

3. “Sequence of Parts/ Tasks”

It describes how all of the parts and/or tasks in a whole test are put together.

4. “Relative Importance of Parts/ Tasks”

It explains how the parts and/or the tasks in the test differ in importance.

5. “Number of Tasks/ Items per Part”

The number of items for each part of the test is stated here.

b) The Test Task Specifications

There are some elements of the test task specifications. Basically, the elements will be influenced by the characteristics of the test and the characteristics of the test-takers. However, there are some basic elements that can be found in all of kind of the test. They are the objectives of the assessment, characteristics of the input and output and time allotment.

i) The Objectives of the Assessment

The objective in this part should be the same with the objective that has been already stated in the designing stage. The objective is stated once more here to make the test developers do not lose sight of the specific purpose which has been decided.

ii) Characteristics of the Input

Discussing the input characteristics in language assessment, the topics have closely related to the input of format and input of language characteristics. Input of format characteristics describe much about the way the input is presented while input of language talk a lot about the nature of language that is used.

(1) Input of Format Characteristics

There are some elements of input format characteristics as proposed by Bachman and Palmer (1996: 52 - 53). They are form, language, length, type, degree of speediness, and vehicle:

1. “Form”

This part describes whether the input is in the form of language, non-language (pictures, gestures, actions, etc) or both.

2. “Language”

If the input in the form of language, it should be clearly explained here whether the input is given in native language, target language or both.

3. “Length”

The length of the input is influenced by the amount of interpretation required. In this part, it should be clear whether the input is given in single words, phrases, sentences, paragraph or extended discourse.

4. “Type”

The type of the input can be in the form of an item or a prompt. As the example, the input in a telephone conversation can be characterized as a series of items. Meanwhile, the input of the request from a department head can be categorized as a prompt.

5. “Degree of Speediness”

It talks about the rate at which the test-takers have to process the information in the input.

6. “Vehicle”

It should be cleared how the input is delivered. The input can be presented live, reproduced (via audio recording) or both.

(2) Input of Language Characteristics

There are two aspects that should be considered as the characteristics of the language input. They are language characteristics and topical characteristics.

1. Language Characteristics

In this part, the input of language which is given is described deeply in terms of the grammatical characteristics (vocabulary, syntax, phonology, etc), textual characteristics (rhetorical, conversational organization, etc), functional characteristics (ideational, manipulative, etc), and sociolinguistics characteristics (dialect, register, cultural awareness, etc).

2. Topical Characteristics

It refers to the type of information in the input. The type can be personal, cultural, academic, or technical.

iii) Characteristics of the Output

Like the characteristics of the input, the output also consists of the output format characteristics and the output of language characteristics. The elements of each aspect are similar to the elements of the input characteristics.

iv) Time Allotment

Different types of test will need different times to complete. It might be need more or less time. Therefore, the test-developer should specify the suitable time allotment for the test-takers to maintain the control of the time needed by the test-takers to finish the test.

c) Criteria of Correctness

The criteria of assessment correctness should be clear enough. The aim is to ensure the scores that are given is meaningful based on the construct that have been developed (Buck, 2001). The test-developers can specify the criteria of the correctness in two aspects. They are the areas of language knowledge and the topical knowledge (Bachman& Palmer, 1996).

Table 6: The Areas of Language Knowledge proposed by Bachman and Palmer (1996: 68)

Organizational Knowledge (how utterances or sentences and texts are organized)	Grammatical Knowledge (how individual utterances or sentences are organized)	knowledge of vocabulary
		knowledge of syntax
		knowledge of phonology/morphology
	Textual Knowledge (how utterances or sentences are organized to form texts)	knowledge of cohesion
		knowledge of rhetorical or conversational organization
Pragmatic Knowledge (how utterances or sentences and texts are related to the communicative goals of the language user and to the features of the language use setting)	Functional Knowledge (how utterances or sentences and texts are related to the communicative goals of language users)	knowledge of ideational functions
		knowledge of manipulative functions
		knowledge of heuristic functions
		knowledge of imaginative functions
	Sociolinguistic Knowledge (how utterances or sentences and texts are related to features of the language use setting)	knowledge of dialects/varieties
		knowledge of registers
		knowledge of natural or idiomatic expressions
		knowledge of cultural references and figures of speech

Topical knowledge is a kind of knowledge that is related to the test-takers real-world knowledge. In order to designing an assessment and specifying its criteria of correctness, the test-developers should pay attention on this aspect. Certain language assessments that presuppose cultural or topical knowledge as a part of certain test-takers might be easier for those who have that knowledge and might be more difficult for those who have not the knowledge (Bachman & Palmer, 1996).

3) The Administration Stage

In the step of test administration, there are some procedures which are related to the collecting of the qualities of the usefulness. It aims to investigate the

developed test and make revision to improve its qualities. This step also consists of the procedures to analyze the test score.

To get feedback about the qualities of the usefulness, it can be obtained through the questionnaires, observation and interviews. The data is collected in two separated times. They are pre-testing and operational testing knowledge (Bachman & Palmer, 1996). The data in pre-testing step is the data that is obtained before the test is actually used. Meanwhile, the data from operational testing is obtained during the implementation of the test as what have been designed before.

In terms of analyzing the test score, Bachman and Palmer (1996: 92) describe some procedures. They are describing test scores, reporting test scores, item analysis, estimating reliability, and investigating validity.

1. "Describing Test Scores"

The test score can be analyzed using descriptive statistics to know the quantitative characteristics of the test scores.

2. "Reporting Test Scores"

In order to report the test scores, statistical procedures can be used. It can be used to determine the most effectively ways in reporting the scores.

3. "Item Analysis"

Analyzing the items of the test tasks should be done through the various statistical procedures.

4. “Estimating Reliability”

Statistical procedures are used to estimate the consistency of the test scores across different specific conditions of the test use.

5. “Investigating the Validity”

To analyze the validity of the test scores, both of the quantitative and qualitative procedures can be used.

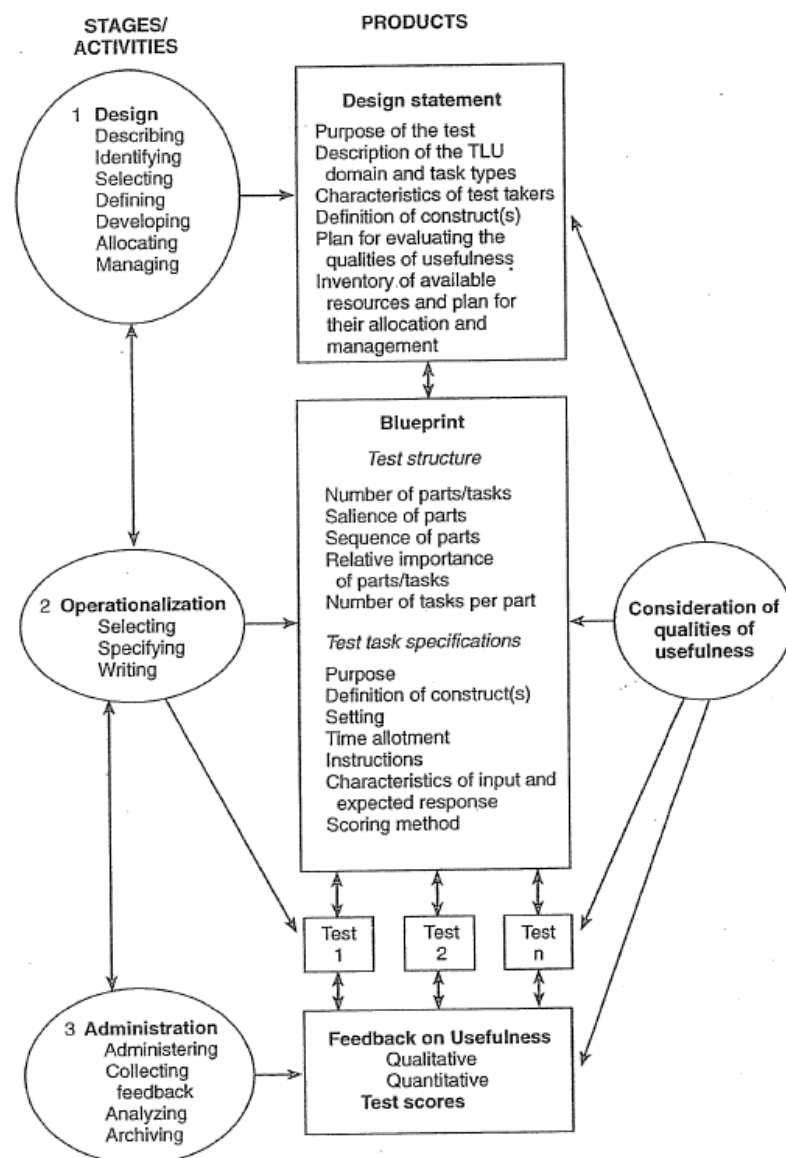


Figure 3: Test Development

3. The Nature of Listening for Young Learners

Listening as one of the four language skills is important for young learners. In this section, there will be some discussions related to the topic. They are teaching listening to young learners and listening assessment for young learners.

a. Teaching Listening to Young Learners

Listening is the receptive skills because it focuses on receiving information from the outside source. However, it does not mean that listening is the passive activity. Learners can be actively engaged in various kinds of listening activities and tasks. The teacher can prepare the children to replicate the sounds when they speak through listening activities (Linse, 2005).

Learners who are usually trained in written language will get difficulties when they faced listening suddenly. The teacher should provide listening regular practices and activities to make them familiar (Rivers, 1981).

Teaching English to young learners may give some advantages for the learners. At least, there are six benefits which are described as follows(Ferguson, et al., 2009: 20):

1. Listening can be used to energize students or to create a quiet atmosphere.
2. As children become accustomed to the teacher's repeated greetings and instructions, they internalize the words and phrases, learn to respond appropriately, and begin to participate in genuine conversations.
3. Recorded listening material provides opportunities for children to hear different accents.
4. Children use the teacher's pronunciation as a model for their own pronunciation.
5. Listening activities develop children's abilities to focus on both general and specific information, raising their awareness that they can understand a story or conversation without knowing the meaning of every word.
6. Audio lessons using cassette tapes and CD's are useful for modelling correct pronunciation and offering a different approach for students to hear spoken English.

According to Brewster and Ellis (2002), designing listening task should be done in three stages. They are pre-listening activities, while-listening activities and post-listening activities. Pre-listening activities are the activities that should be done as a preparation for listening. Teacher can introduce some of useful words, structures and vocabulary for the children together with providing motivation and interest for them. During the listening, children are also provided with the activities. This is called as while-listening activities. After listening, the teacher usually asks the learners to do some activities like drawing or sticking something or answer some questions. This is post-listening activities.

In order to design listening activities, the teacher can use the young learners' characteristics as the basic information to develop activities that suit for them. Songs, stories and games can be used as the materials which are developed in various activities. Brewster and Ellis (2002: 102-104) illustrate some of the listening activities as listed below:

1. listen and repeat
2. listen and discriminate
3. listen and perform actions/ follow instructions
4. listen and draw/ colour
5. listen and predict
6. listen and guess
7. listen and label
8. listen and match
9. listen and sequence
10. listen and classify
11. listen and transfer information

b. Listening Assessment for Young Learners

As the integral part of teaching and learning process, listening assessment should be done as the part of the teaching and learning listening for young learners. Sometimes, listening skill are assessed together with the speaking skill.

However, it is still possible to assess listening skill only because listening is very important not only in language learning but also in learning itself (McKay, 2006). Learners need to understand the teacher's talk and any language input which is given. Moreover, listening is important for young learners to access new language and find out how the language works (Georgiou & Pavlou, 2003).

Listening as the receptive skill is quite hard to be assessed because the listening performance itself is invisible. The assessment developer needs to provide the listening assessment activities which are observable. It means that the performances are able to be seen and heard.

Table 7: Observable Performance of the Four Skills proposed by Brown (2004: 118)

	Can the teacher directly observe. . .	
	the process?	the product?
Listening	No	No
Speaking	Yes	No*
Reading	No	No
Writing	Yes	Yes

*Except in the case of an audio or video recording that preserves the output.

Listening assessment can give evidence that the learners are start to explore their limited knowledge to interpret the meaning of what they hear. Young learners need time to produce the language. It is called as 'silent period'. Before they are ready to produce the language, they process the language and explore how the language works. In this case, the learners do not always need to respond verbally (Georgiou & Pavlou, 2003). There some ways and techniques to get responses from the learners in form of non-verbal responses. Brewster and Ellis (2002: 254) are proposed some assessment activities in order to assess listening skill. They are:

1. listen and discriminate between sounds
2. listen and point to things or follow instructions
3. listen and select the appropriate pictures
4. listen to a description and draw or colour a picture
5. listen to a description and label the picture
6. listen and match two pictures or a word and a picture
7. listen and sequence pictures, words or sentences
8. listen to a description or story and tick items on a simple chart
9. listen and take simple notes
10. listen and complete gaps in words or sentences
11. listen to a narrative and trace a route on a map or plan
12. listen and select the correct response (multiple choice)

Rivers (1981: 166) also has similar opinion that the most effective listening skill assessment is when it is designed with a purposeful context. The examples can be seen follows:

1. Students listen for instructions which must be carried out in some practical ways.
2. Students listen to subject matter which must be learned and which itself will be tested.
3. Students listen as a part act of communication, as a part of a conversational interchange which can only continue when there is comprehension.

There are some of the studies which concern on assessing young language learners. Espinosa and Lopez (2007) discuss the issues related to the assessing young learners. One of the points is the measurement strategies for the students. Meanwhile, Shephard (1994) had already talked about the challenges to assess the young learners appropriately in few years ago. It is important to know the children's needs and to make the assessment which fulfills the criteria of the appropriate listening assessment.

B. Conceptual Framework

Listening assessment as the integral part of teaching and learning English is important to measure the learners' achievement which can support the learners'

abilities in other language skills. However, listening assessment which can evaluate the learners' ability accurately should pay attention to some criteria.

The ideal listening assessment for young learners should be considered based on the two big aspects. First, the assessment developer has to know the young learners' characteristics and their needs. Since young learners are different from adult, they have their own uniqueness and characteristics. Those can be used as the basic information to develop the listening assessment that is suitable for them. Second, the assessment developer also needs to pay attention on the principles of the assessment. Those principles are used as the indicators of well-made assessment.

However, listening assessment for the fifth grade students in SDN Rejodani does not fulfil those criteria above. Based on the classroom observation that has already stated, the listening assessment process is inappropriate in terms of the content, the form and the methods. Therefore, it is necessary to conduct a research on designing the ideal listening assessment for the fifth grade students of SDN Rejodani.

In order to produce the ideal listening assessment for them, there are some steps which should be done. The steps are expected to improve the qualities of the assessment based on the principles of the ideal assessment. Bachman and Palmer (1996) proposed some of the steps to design the test which can be adapted to design a listening assessment.

First, it is important to conduct need analysis. The data can be collected through observation, interview and questionnaires. The information from the data is used to find out the learners' characteristics and needs.

Second is the process of designing the developed listening assessment. The assessment developer needs to describe the relations between the assessment tasks and assessment scores with the language use first in designing stage. Then, the blueprint of the assessment is made based on the result of learners' need analysis, standard of graduate competence, standard of competence and basic competence. The components of the blueprint of the assessment are the test structure, the test task specifications and criteria of correctness. The actual assessment task is made based on the blueprint. The first draft of the listening assessment is validated through the expert judgement before it is implemented.

The follow-up stage after the implementation is the administration stage. The assessment developer evaluates the draft of listening assessment after it is implemented. It aims to improve the qualities of the listening assessment itself. The information to revise the listening assessment is get from the respondents, the teachers and the learners. It can be obtained through interview and questionnaires. The assessment scores should also be analyzed based the criteria of correctness which has already been stated in the blueprint of the assessment. After completing all of the steps above, the availability of a set of the ideal listening assessment for the fifth grade students of SDN Rejodani can be produced. The diagram of the conceptual framework that has been described can be seen on the next page:

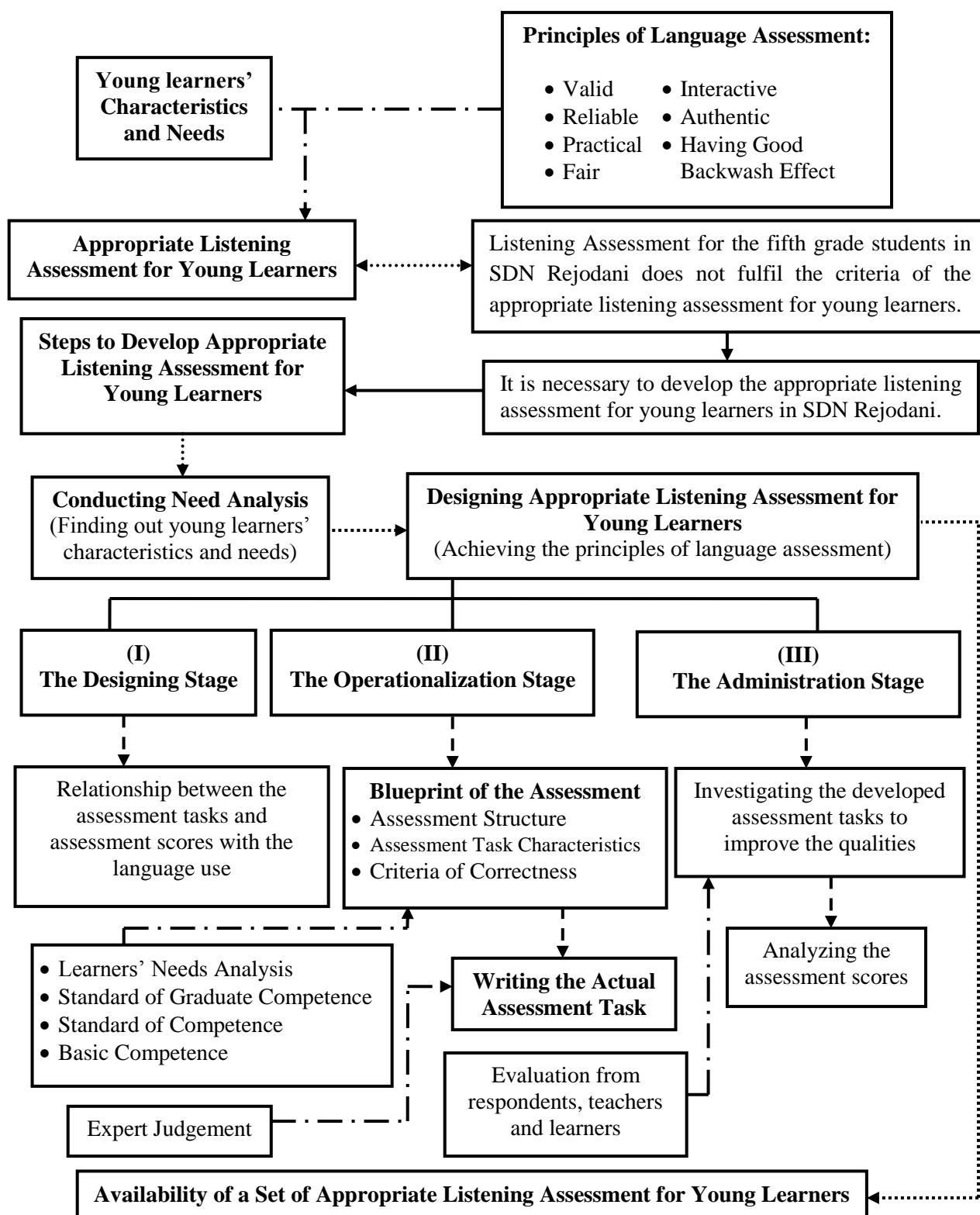


Figure 4: Conceptual Framework

CHAPTER III

RESEARCH METHOD

B. Research Design

The research was categorized as Educational Research and Development (R&D) study. The kind of research study was used to develop and validate the educational product. The examples of the products were developed materials, media, management system, etc. The development was based on the detailed of specification which derived from the data of specific needs. The product of the research should be tried out and revised in order to achieve the effectiveness of the product (Borg & Gall, 1983).

In the research study, I tried to design a set of listening assessment for the fifth grade students of SDN Rejodani Sleman Yogyakarta. First, I identified the learners' need analysis by distributing questionnaires for them. After identifying the need analysis, a blueprint of the listening assessment was made and it was developed to get the first draft. The first draft had been revised and validated by the experts to get a quality assurance before the implementation. After it was revised, the second draft was made. It was tried-out twice. The first try-out aimed at collecting the data of the teacher, the observer and students' evaluations while the second try-out aimed at collecting the second assessment scores to analyze the reliability of the designed assessment. The second draft was revised and validated again to get the final product of the designed listening assessment.

C. Research Subjects

The research subjects were the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta. There were 20 male and 21 female students in the class. Their age were varied from ten to eleven years old and came from different social background. Most of their parents worked as farmers, teachers and employees.

D. Research Setting

The study was done in SDN Rejodani. It is located on Ngaglik Sleman Yogyakarta. The school has several classrooms, a mushola, a canteen, a teacher's room, a headmaster's room and some toilets for boys and girls.

E. Research Instruments

There were three kinds of instruments that were used in the research. They were questionnaires, the observation sheet and interviews. The instruments were used to get the data about the learners' need and the data to revise the designed listening assessment.

1. Questionnaires

There were two types of the questionnaires. First, I designed the questionnaire to collect the data related to the learners' need. The organization of the questionnaire is presented as follows:

Table 8: Blueprint of the Need Analysis Questionnaire

The Purpose of the Questions	Part	Question Numbers	References
To get some information related to students' profile	A		Masuhara via Tomlinson (1998: 240)
To get some information related to students' internal factors in assessment process	B		Nunan (2004:120) Georgiou & Pavlou (2003: 8)
To get some information related			Georgiou & Pavlou

to the assessment process in the classroom			(2003: 21)
To get some information related to language input for young language learners			Pinter (2006: 30)
To get some information related to the difficulty aspects in listening skill			Pinter (2006:45)
To get some information related to the principles underlying the assessment and evaluation			Ontario (2011:29)

The second type of the questionnaire was the evaluation questionnaire. It was used to collect the information from the expert and the learners. The data were used to revise the designed assessment.

The evaluation questionnaire for the expert was given after the first draft of the designed listening assessment was made. The result was functioned as the information to revise the designed assessment before it was tried out.

Table 9: Blueprint of the Evaluation Questionnaire administered to the Experts

The Purpose of the Questions	Part	Question Number	References
To find some information related to the respondents' profile	A		Tomlinson (1998: 240)
To find some information to evaluate the correlation between the designed assessment and the national curriculum	B		Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar (2008: 11, 17-19)
To find some information to evaluate validity of the designed assessment			Bachman& Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation& Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment			Gordon (2007: 202); Nation& Newton (2009: 167); Brown (2004: 31-32)
To find some information to evaluate practicality of the designed assessment			Bachman& Palmer (1996: 155); Baxter (1997: 28); Nation& Newton (2009: 168); Brown (2004: 31)
To find some information to evaluate the authenticity and the			Bachman& Palmer (1996: 151-152); Brown (2004: 35)

interactiveness of the designed assessment			
To find some information to evaluate backwash effect of the designed assessment			Bachman& Palmer (1996: 153-155); Baxter (1997: 28-29)
To find some information to evaluate the fairness of the designed assessment			Cameron (2001: 226); Wakeford (2003: 45-46)

Another evaluation questionnaire was distributed to the students after the designed assessment was tried-out. The result was used to design the final draft of the designed listening assessment. The organization of the evaluation questionnaire is presented in the table below.

Table 10: Blueprint of the Evaluation Questionnaire administered to Students

The Purpose of the Questions	Question Number	References
To find some information related to respondents' profile		Tomlinson (1998: 240)
To get some information related to the students attitude toward the designed assessment		Georgiou & Pavlou (2003:4-6, 8)
To find some information related to the validity of the designed assessment.		Bachman& Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation& Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment		Gordon (2007: 202); Nation& Newton (2009: 167); Brown (2004: 31-32)
To find some information related to the practicality of the designed assessment		Bachman& Palmer (1996: 155); Baxter (1997: 28); Nation& Newton (2009: 168); Brown (2004: 31)
To find some information related to the authenticity and the interactiveness of the designed assessment		Bachman& Palmer (1996: 151-152); Brown (2004: 35)
To find some information related to the backwash effect of the designed assessment		Bachman& Palmer (1996: 153-155); Baxter (1997: 28-29)

2. Observation Sheet

Another instrument in the research was observation sheet which was given to the teacher and the observer. It was done during the try-out of the designed listening assessment. The observation sheet is organized as follows:

Table 11: Blueprint of the Classroom Observation Sheet

The Purpose of the Questions	Question Number	References
To find some information related to respondents' profile		Tomlinson (1998: 240)
To get some information related to the students attitude toward the designed assessment		Georgiou & Pavlou (2003:4-6, 8)
To find some information related to the correlation between the designed assessment and the national curriculum		Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar (2008: 11, 17-19)
To find some information to evaluate validity of the designed assessment		Bachman& Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation& Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment		Gordon (2007: 202); Nation& Newton (2009: 167); Brown (2004: 31-32)
To find some information related to the practicality of the designed assessment		Bachman& Palmer (1996: 155); Baxter (1997: 28); Nation& Newton (2009: 168); Brown (2004: 31)
To find some information related to the authenticity and the interactiveness of the designed assessment		Bachman& Palmer (1996: 151-152); Brown (2004: 35)
To find some information related to the backwash effect of the designed assessment		Bachman& Palmer (1996: 153-155); Baxter (1997: 28-29)
To find some information related to the fairness of the designed assessment		Cameron (2001: 226); Wakeford (2003: 45-46)

3. Interviews

I interviewed the teacher and students after the first try-out of the second draft in order to get their opinion related to the designed assessment. The result was used to revise the second draft of the designed assessment into the final draft.

F. Validity and Reliability of the Instrument

Validity and reliability were very important in the research instruments. The instruments were valid if it really measured what supposed to be measured. Meanwhile, reliability of the research instruments was important to enhance the trustworthiness of the data and to reduce the subjectivity in analyzing the data.

There were three kinds of the research instruments that were used in the research study. The description of validity and reliability for each type of the research instruments were listed below.

1. Questionnaires

There were two kinds of questionnaires in the research study. They were need analysis and evaluation questionnaires. Both of them were designed by considering content validity. It means that the items in the questionnaires were written in line with the blueprint of the questionnaires. The aim was to make sure that the questionnaires were really designed based on the theory as proposed by some experts. To enhance the reliability of the questionnaires, the Cronbach Alpha's statistics was used.

2. Observation Sheet

The observation sheet was made based on the observation's blueprint. The content validity was used to examine whether the observation sheet represented the theories that was used to find out what was happening in the classroom or not. Meanwhile, enhancing reliability of the observation sheet was done by considering the external audit.

3. Interviews

The framework of the interviews was established based on the blueprint of the evaluation questionnaire and the observation sheet. The respondents were given certain questions related to the framework but they were free to talk more and gave their own opinion related to the framework. To enhance the reliability of the interviews, the external audit was used in coding the responses.

G. Data Collection Techniques

The data in the research were collected in three separated times. First, I collected the data about the learners' needs as the information to develop the blueprint of the assessment. Second, the evaluation questionnaire for the expert was given after the designed listening assessment was made before the try-out. The result was used to revise and prepare the product to be tried-out. Third, the evaluation questionnaires were distributed to the students and the assessment scores also were analyzed after the first try-out. The data from the classroom observation in the form of observation sheet also were got from the teacher and the observer during the first try-out of the assessment. In the second try-out of the second draft, the assessment scores were also collected once again and were analyzed together with the first one to get the data of the reliability designed assessment.

H. Data Analysis Techniques

The data of the research study were analyzed in two ways. First, the data from the observation sheet and the interviews were analyzed qualitatively. The data

from classroom observation were reported in the form of field notes and the observation sheet. The data from the interview with the teacher and students were collected after the try-out of each assessment topic so do the data from the observation sheet.

Second, the data from the evaluation questionnaire were analyzed quantitatively. The information was used to revise and design the listening assessment. The data were analyzed by using Likert Scale. Likert Scale asked respondents to indicate the strength of agreement or disagreement with a given statement on five or seven point range. However, respondents tended to choose 'neutral' statement when they were provided by five choices of point ranges since it was the easiest alternatives which did not demand them to think more (Arikunto, 2006: 241). Therefore, I used the version of Likert Scale from Arikunto. Four pointsrange was implemented in Likert Scale by Arikunto. They were:

4 = SA, if the respondents strongly agree with the statement

3 = A, if the respondents agree with the statement

2 = D, if the respondents disagree with the statement

1 = SD, if the respondents strongly disagree with the statement

After the quantitative data were collected, the central tendency measure was used by calculating the mean value for each of the assessment elements. It was calculated by following formula as proposed by Suharto (2005: 58).

$$Mn = \frac{\sum fx}{N}$$

Figure 5: The Mean Formula proposed by Suharto (2005: 58)

All of the mean values were put into the category. The category was decided by calculating the range which was used to classify the mean values in classes. To calculate the range of each class, the formula below was used as what proposed by Suharto (2005: 59).

$$R = Xt - Xr$$

Figure 6: The Range Formula proposed by Suharto (2005: 59)

After the range of the class was found, the result was divided by the desired number of categories. The research study used five categories to classify the mean values. Based on the calculation, the class interval could be represented as follows:

Table 12: Quantitative Data Conversion

Scale	Category	Interval of Mean
5	Very good	3.41 – 4.01
4	Good	2.83 – 3.40
3	Fair	2.22 – 2.82
2	Poor	1.61 – 2.21
1	Very Poor	1.00 – 1.60

I. Research Procedure

The research procedure in the research study was adapted from Borg and Gall's model (Borg& Gall, 1983: 775). There were ten steps which were promoted by Borg and Gall to develop educational products. They were research and information collecting, planning, development of the preliminary form of the product, preliminary field testing, main product revision, main field testing, operational product revision, operational field testing, final product revision and dissemination and implementation.

I modified the model so that it was suitable for the research study. The steps were described as follows:

1. Conducting Needs Analysis

To get the information about the learners' need, I conducted need analysis in the form of questionnaire. The result was used as a basic to develop the blueprint of the listening assessment.

2. Writing the Blueprint of the Listening Assessment

The blueprint of the listening assessment was able to figure out the kind of listening assessment as whole. It was made based on the result of learners' need analysis. The standard of graduate competence, standard of competence and basic competence in national curriculum were also considered as a standard to develop the listening assessment.

3. Designing the First Draft of the Listening Assessment

The first draft of the listening assessment was made based on the blueprint. The product was not only the assessment sheet but also the audio recording.

4. Getting Expert Judgement

The listening assessment should be validated by the expert before it was tried-out. The evaluation of the first draft was done by giving the evaluation questionnaire to the expert.

5. Revising the First Draft of the Listening Assessment

Any feedback from the expert through the evaluation questionnaire was analyzed quantitatively. The result was used to evaluate the first draft.

6. Designing the Second Draft of the Listening Assessment

The result of the evaluation of the first draft was used as a basic to design the second draft. It was prepared for the try-out to get suggestions from the teacher, students and the observer.

7. Trying-Out the Second Draft of the Listening Assessment

The second draft of the listening assessment was tried-out twice in the classroom. During the first try-out, classroom observation was also recorded in the form of field notes and in the observation sheet. The evaluation questionnaire for the students was given after each of the assessment' topic was tried-out. I also conducted the interviews with the teacher and students. The second try-out was done to get the second assessment scores. It was analyzed together with the assessment scores in the first try-out to get the reliability of the designed assessment.

8. Revising the Second Draft of the Listening Assessment

The classroom observation data, the result of the evaluation questionnaire and the interview transcripts were analyzed to get the information related to the

second draft of the designed listening assessment. The data were used to revise the second draft.

9. Writing the Final Draft of the Listening Assessment

The second draft of the listening assessment was revised in all aspects that need to be revised. Finally, the final draft could be made.

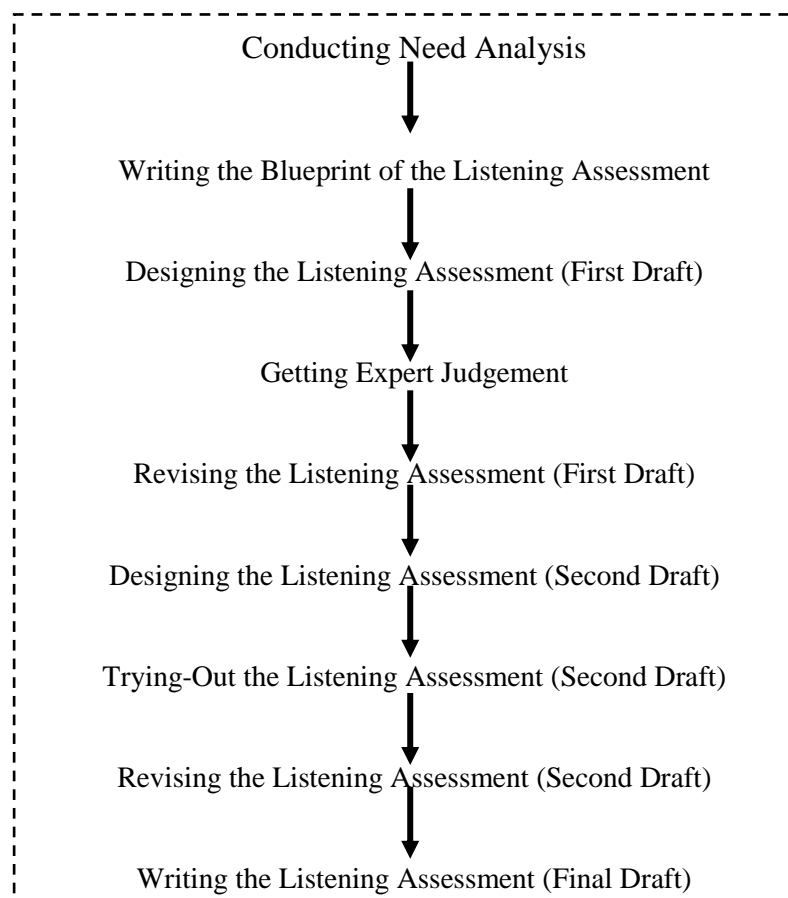


Figure 7: Research Procedure

CHAPTER IV

RESEARCH FINDINGS AND DISCUSSIONS

J. Research Findings

There were nine steps which had been conducted to design the appropriate listening assessment in the research study. The figure can be seen below.

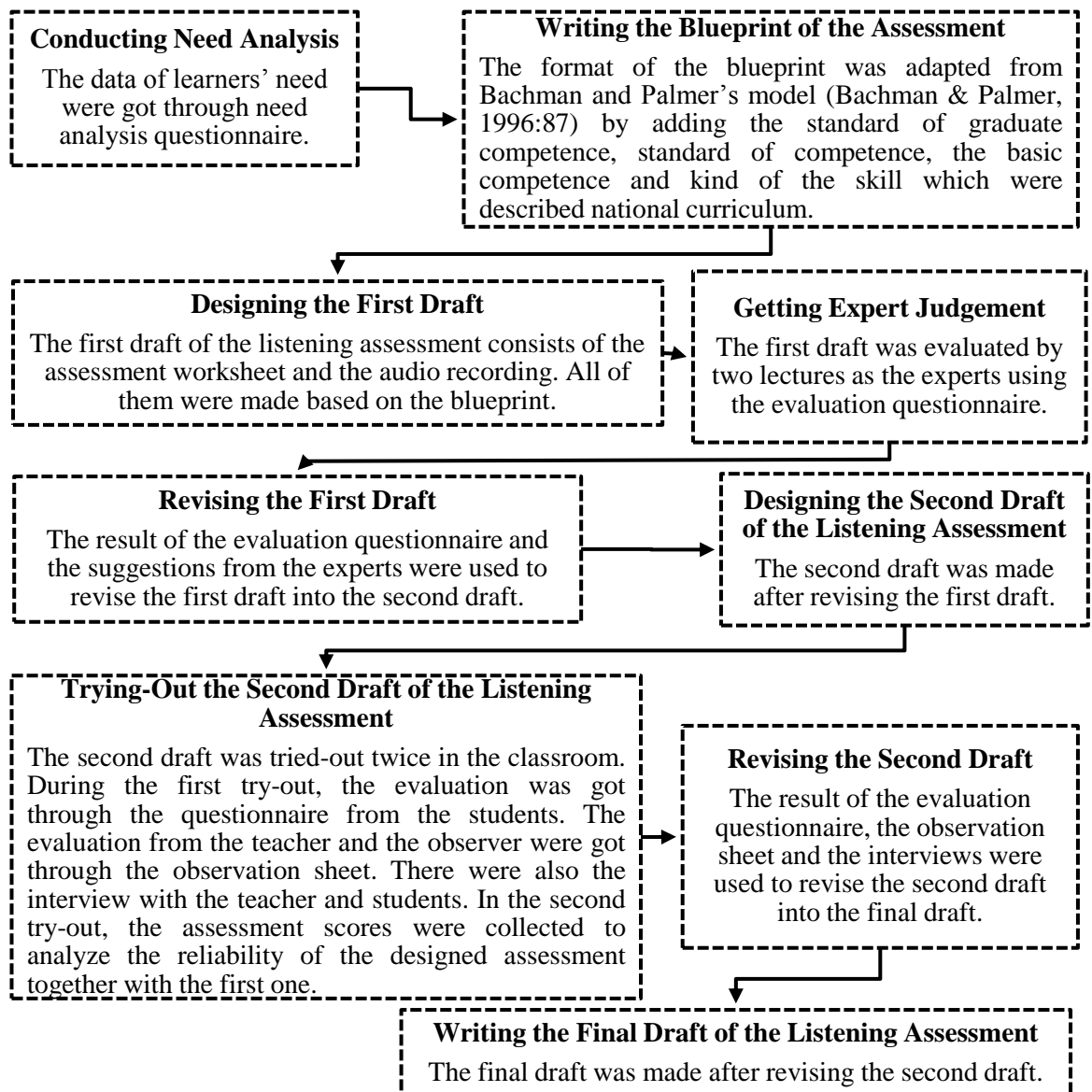


Figure 8: Steps in Developing the Appropriate Listening Assessment

1. Designed Listening Assessment

a. The Result of Learners' Needs Analysis

The first data of the students' need were got through the need analysis questionnaire. The questionnaire consisted of two parts and 25 items which were developed from some elements proposed by the experts. It fulfilled the content validity because the items were written in line with the elements which were already designed in the blueprint.

Table 13: Blueprint of the Need Analysis Questionnaire

The Purpose of the Questions	Part	Question Numbers	References
To get some information related to students' profile	A	-	Masuhara via Tomlinson (1998: 240)
To get some information related to students' internal factors in assessment process	B	1, 2, 3, 6, 11, 24, 25	Nunan (2004:120) Georgiou &Pavlou (2003: 8)
To get some information related to the assessment process in the classroom		10, 15, 17, 19, 20, 21, 22, 23	Georgiou &Pavlou (2003: 21)
To get some information related to language input for young language learners		7, 12, 13, 14	Pinter (2006: 30)
To get some information related to the difficulty aspects in listening skill		9, 16	Pinter (2006:45)
To get some information related to the principles underlying the assessment and evaluation		9, 18, 5, 4	Ontario (2011:29)

It is also regarded reliable since the Alpha's coefficient was 0.989. The data computation details can be checked in Appendices while the further description about the result of the questionnaire can be seen below.

The first part was designed to get the information about the students' profile. There were 41 students in the classroom that came from different social and

economic background. Their parents worked as farmers, teachers, employees, etc.

The data of the students' profile can be seen in the following table.

Table 14: Data of the Fifth Grade Students of SDN Rejodani

Number of Students	Sex		Age
	Male	Female	
41 Students	20	21	10-11

The second part consisted of five elements which were aimed at collecting information related to the listening assessment in the classroom. Those were the students' internal factors in assessment process, the assessment process in the classroom, language input for young learners, the difficulty aspects in listening skill, and the principles underlying the assessment and evaluation.

The first element was the students' internal factors in assessment process. Based on the result of the questionnaire, most of the students loved to learn English and agreed that English was easy. They were also optimistic to understand the expressions in English more. To enhance their abilities in English lesson, students still learnt English at home although the teacher did not give assignments. Although they were afraid of the English assessment because of the bad marks, most of them still loved to do that.

Table 15: Students' Responses on Need Analysis Questionnaires of the First Element by Question Number, Learners' Percentage and Category

Elements of the Items	Items	Question Number	Learners' Percentage	Category
students' internal factors in assessment process	I love to learn English.	1	44%	Agree
	English is easy.	2	44%	Agree
	I want to understand the expressions in English.	3	73%	Agree
	I learn English although the teacher doesn't give assignments.	6	39%	Agree
	I am not afraid of the English assessment.	11	41%	Disagree

	The students love to do the English assessment.	24	46%	Agree
	The students are afraid of the bad mark in the English assessment.	25	39%	Strongly Agree

The second element was the assessment process in the classroom. The data of the second elements showed that the teacher was always reviewing the lesson before assessing the students. The teacher also gave clear instructions before they did the test and always tried to help the students when they found any difficulties related to the test. However, they still considered that the English test made by the teacher is not fun. They also agreed that the items in the test were not varied. Feedback was also important in language assessment. In the research study, most of the students agreed that the teacher gave some feedback to them by discussing the test together after it was finished. Related to the layout, I wanted to know the layout of the assessment that the students like most. Based on the result of the questionnaire, they loved the English listening assessment which has colourful pictures.

Table 16: Students' Responses on Need Analysis Questionnaires of the Second Element by Question Number, Learners' Percentage and Category

Elements of the Items	Items	Question Number	Learners' Percentage	Category
Assessment process in the classroom	The teacher always gives the review of the lesson before giving the test.	10	76%	Agree
	The teacher always explains how to do the test before it starts.	15	39%	Agree
	My friends, the teacher and I discuss the English assessment together after it finished.	17	51%	Agree
	The English assessment which is made by the teacher is fun.	19	41%	Disagree
	The teacher gives the various	20	39%	Disagree

	test items.			
	I love the English assessment which has some pictures.	21	44%	Agree
	I love the colourful English assessment.	22	56%	Agree
	The teacher helps me when I find any difficulties in the test items.	23	68%	Agree

The third element was the language input for young learners. The input could be in the form of songs, stories and daily routines in the classroom. I wanted to know how much the input that the teacher gave to the students through the element. Based on the result of the questionnaire, most of the students agreed that the teacher gave the input of language in the form of daily instructions and simple stories. Their teacher also often tried to introduce the new English songs and they loved it.

Table 17: Students' Responses on Need Analysis Questionnaires of the Third Element by Question Number, Learners' Percentage and Category

Elements of the Items	Items	Question Number	Learners' Percentage	Category
Language input for young language learners	I like the English songs.	7	36%	Agree
	The teacher often introduces the English songs.	12	36%	Agree
	The teacher tells the story in English.	13	46%	Agree
	The teacher gives the instructions in English.	14	49%	Agree

The fourth element was the difficulty aspects in listening skill. Based on the result of the questionnaire, most of them understood their teacher's expressions in English. The teacher also helped the students when they found the problems related to the difficult words in listening skill.

Table 18: Students' Responses on Need Analysis Questionnaires of the Fourth Element by Question Number, Learners' Percentage and Category

Elements of the Items	Items	Question Number	Learners' Percentage	Category
The difficulty aspects in listening skill	I understand the teacher's expressions in English.	8	41%	Strongly Agree
	The teacher explains the difficult words.	16	66%	Agree

The fifth element was the principles underlying the assessment and evaluation.

Most of the students learned English since they were in Kindergarten. When they were in Elementary School, their teacher often gave the English assessment in the classroom. However, they could not do the assessment well when they were sick and tired. Parents as one of the supporting factors in the assessment did not influence the students because most of them stated that their parents did not help them to learn English at home.

Table 19: Students' Responses on Need Analysis Questionnaires of the Fifth Element by Question Number, Learners' Percentage and Category

Elements of the Items	Items	Question Number	Learners' Percentage	Category
the principles underlying the assessment and evaluation	The teacher often gives the English test.	9	73%	Agree
	I can't do the test well when I am tired and sick.	18	46%	Strongly Agree
	I have learned English since I were in Kindergarten.	5	73%	Agree
	My parents help me to learn English.	4	39%	Disagree

Another source of the students' need data were from the observation. Based on the observation, I found that the content, the form, and the methods of the listening assessment were inappropriate. The listening assessment also did not fulfil the principles of the assessment which were essential to measure the students' competence accurately.

b. Blueprint of the Listening Assessment



The blueprint of the listening assessment made based on the students' needs analysis and the observation. Its format was adapted from Bachman and Palmer's model (1996: 87) which consisted of three sections. They were the assessment structure, the assessment task characteristics, and criteria of marking.

1) The Assessment Structure

The assessment structure gave brief descriptions on how the assessment was arranged. The section consisted of different number of parts, sequence of parts, and the number of items for each assessment topic. There was the detail information related to the number of parts including the instructions for each part of the designed assessments. In this section, how the parts were arranged and number of items for each part also was clearly informed.

The first assessment was *In the Classroom*. The topic was chosen since the materials were classroom instructions. It was suitable with the standard of competence for the fifth grade elementary school students. The assessment consisted of one part which asked the students to match the classroom instructions that they heard with the correct pictures. Because there was only one part, there was no sequence of parts in the first assessment. The assessment also consisted of ten items.






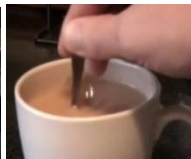


Table 20: The Assessment Structure of the First Assessment

First Assessment “In the Classroom”	
Number of parts	<p>One part ↓ Match the classroom instructions in audio recording with the correct picture</p> <p>ex:</p> <div style="display: flex; align-items: center; justify-content: center;">  <div style="border: 1px solid black; padding: 5px; margin-left: 20px;">✓</div> </div> <p> Clean the board</p>
Sequence of parts	-
Number of items/ part	Ten (10)

The topic of the second assessment was *A Cup of Tea*. It was chosen by considering the standard of competence of the fifth grade elementary school students. I choose *A Cup of Tea* as a topic because the students were familiar with the topic. The instructions in the topic were also simple and suitable with their language level.

There were two parts in the second assessment. First, the students had to match the picture of the materials about making a cup of tea with the audio recording that they heard. Second, they heard the steps of making a cup of tea then they had to rearrange the pictures based on what they heard. Because, there was more than one part, the sequence of the part was that the students did the part A first then part B. Part A consisted of five items while part B consisted of seven items.





Table 21: The Assessment Structure of the Second Assessment

Second Assessment “A Cup of Tea”	
Number of parts	Two parts (Part A and Part B) ↓ Part A : Put the number on the right picture based on materials and ingredients about making a cup of tea in the audio recording ex: <div style="display: flex; align-items: center; justify-content: center;">     a kettle </div> <div style="display: flex; justify-content: center; gap: 50px; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px; text-align: center;">1</div> <div style="border: 1px solid black; width: 30px; height: 30px; margin: 0 auto;"></div> <div style="border: 1px solid black; width: 30px; height: 30px; margin: 0 auto;"></div> </div>
	Part B : Arrange the pictures based on the steps of making a cup of milk tea in audio recording ex: <div style="display: flex; align-items: center; justify-content: center;">    </div> <div style="display: flex; justify-content: center; gap: 50px; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px; text-align: center;">3</div> <div style="border: 1px solid black; padding: 2px 10px; text-align: center;">5</div> <div style="border: 1px solid black; padding: 2px 10px; text-align: center;">1</div> </div>
	 First, put some water in the kettle. Second, boil the water. Third, add some sugar in the cup. Fourth, put the teabag in the cup.
	Sequence of parts Test-takers do the part A first and continue to the part B.
Number of items/part	Part A: Five (5) Part B: Seven(7)

The last topic of the assessment was *Good Daily Activities*. The topic was chosen based on the standard of graduate competence in National Curriculum. It was also chose based on the result of the observation that the teacher ever introduced them with the simple story. The third assessment assessed the students' competence in understanding a simple story.

The story only consisted of one part that asked the students to arrange the pictures based on the story they heard. There was also no sequence of parts since it only consisted of one part. There were ten items in the third assessment.

Table 22: The Assessment Structure of the Third Assessment

Third Assessment “Good Daily Activities”	
Number of parts	<p style="text-align: center;">One part ↓</p> <p style="text-align: center;">Put the number on the right picture in the story based on the good daily activities in the audio recording</p> <p>ex:</p> <div style="display: flex; justify-content: space-around; align-items: center;">     </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 5px;"> <div style="border: 1px solid black; padding: 2px 10px;">2</div> <div style="border: 1px solid black; padding: 2px 10px;">3</div> <div style="border: 1px solid black; padding: 2px 10px;">1</div> <div style="border: 1px solid black; padding: 2px 10px;">4</div> </div> <p>🔊 After I get up, I take a bath and get dressed. Then, I have breakfast. I brush my teeth after having breakfast.</p>
Sequence of parts	-
Number of items/part	Eleven (11)

2) The Assessment Task Characteristics

The assessment task characteristics gave detail information on how the assessment was constructed. The assessment objectives, the input and output characteristics of the assessment and the time which was needed by the students for each assessment topic was specified in the assessment task characteristics.

In the assessment objectives, it was clearly stated the purpose of the assessment including how the materials were assessed. The other parts of the assessment task characteristics were the input and output characteristics.

The input characteristics were divided into the input of format and the input of language characteristics. The input of format covered the form of the input whether it was in the form of language or non-language like pictures, gestures, actions, etc. It was also needed to state in what language the input was given, the length and how it would be presented. Meanwhile, the input of language gave the information on what aspects of language that the designed assessments assessed. It also described kind of the information which was given in the input. It could be personal, academic, or technical. Therefore, the item writers were able to decide whether the vocabulary was restricted in one aspect or not.

The standard of graduate competence of elementary school students, standard of competence and basic competence of the fifth grade students of elementary school and the kind of language skill which were stated in National Curriculum also added into this section as what have been listed below.

Table 23: The Standard of Graduate Competence, Standard of Competence and Basic Competence of the Fifth Grade Elementary School Students and the Language Skill in National Curriculum (2008: 11, 17-19)

Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah
Language Skill	Listening

The first assessment aimed to assess the students' ability to identify the simple classroom instructions that they heard by labelling the correct pictures. There were two elements in characteristics of the input which were the input of format characteristics and input of language characteristics. The input of format in the

first assessment was in the form of language which was classroom instructions and non-language input which was some pictures. The classroom instructions were in the form of audio recording which was in English and the length was relatively short considering the students' level of language. The input of language characteristics were divided into the language characteristics and topical characteristics. The input of language was concerned on vocabulary, syntax and the function of language related to the classroom instructions. Because the topic had been already decided, the topical characteristics were restricted only on the topic.

Characteristics of the output consisted of the output of format characteristics and the output of language characteristics. The output format of the first assessment was in the form of ticking the right pictures based on the audio recording. Because the output format in the form of non-language, the output of language characteristics could not be described. The duration of the assessment was approximately 30 minutes.

Table 24: The Assessment Task Characteristics of the First Assessment

First Assessment “In the Classroom”		
Assessment Objective		This designed assessment assesses students' ability to identify the simple classroom instructions by labelling the correct pictures based on the audio recording about the classroom instructions.
Characteristics of the Input	Input of Format Characteristics	<ul style="list-style-type: none"> - Form: language (simple classroom instructions and non-language (pictures)) - Language: English (Target Language) - Length: relatively short (sentences) - Type: Audio recording about classroom instructions and pictorial input for interpretation - Degree of Speediness: relatively slow - Vehicle: reproduced (via audio recording)

	Input of Language Characteristics	1. Language Characteristics <ul style="list-style-type: none"> - vocabulary: topical vocabulary related to the classroom instructions - syntax: simple sentences and imperatives - function: to identify simple classroom instructions 2. Topical Characteristics: restricted (classroom instructions)
Characteristics of the Output	Output of Format Characteristics	- Form: non-language (labelling correct picture by giving a tick (√))
	Output of Language Characteristics	—
Duration of Tasks		± 30 minutes

The second assessment had two objectives. First, it aimed to assess the students' competence on some vocabularies related to the ways of making a cup of tea. They were asked to label the correct picture based of what they heard. Second, it aimed to assess students' competence to identify some of simple imperative statements related to the topic. They had to arrange the pictures based on what they heard by giving number to each correct picture.

In terms of the characteristics of the input format, the input is in the form of language which was the audio recording about the materials and steps of making a cup of tea and in the form of non-language which was some pictures. The language input is in English as the target language the length of the audio recording was short to medium because the students heard several sentences for whole of the pictures in each part. However, the audio speediness was relatively slow considering students' language level. The input of language concerned on the vocabulary which were related to the materials and steps of making a cup of tea, language syntax which was related to the simple imperative statements and

the language function which was related to the procedure text. All of them were restricted because the input of language limited into one topic only.

The output of format characteristics was in the form of non-language. In part A, the students labelled the correct picture by giving a tick while in part B, they had to arrange the pictures by giving numbers to the correct pictures. The time was set \pm 35 minutes.

Table 25: The Assessment Task Characteristics of the Second Assessment

Second Assessment “A Cup of Tea”		
Assessment Objective		<p>This designed assessment assess students’ ability to:</p> <ol style="list-style-type: none"> 1. number the right pictures based on the materials and ingredients which are needed to make a cup of tea in audio recording 2. identify simple imperative statements by arranging and giving numbers to the pictures based on the steps of making a cup of tea in the audio recording.
Characteristics of the Input	Input of Format Characteristics	<ul style="list-style-type: none"> - Form: language (materials and steps in making a cup of tea) and non-language (pictures) - Language: English (Target Language) - Length: Short to medium (several sentences for whole pictures) - Type: Audio recording about the ways to make a cup of tea and pictorial input for interpretation - Degree of Speediness: relatively slow - Vehicle: reproduced (via audio recording)
	Input of Language Characteristics	<ol style="list-style-type: none"> 1. Language Characteristics <ul style="list-style-type: none"> - vocabulary: topical vocabulary related to the materials and steps of making a cup of tea - syntax: simple sentences and imperatives - function: to identify simple procedure texts (materials and steps of making a cup of tea) 2. Topical Characteristics: restricted (materials and the steps of making a cup of tea)
Characteristics of the Output	Output of Format Characteristics	<ul style="list-style-type: none"> - Form: <ul style="list-style-type: none"> • Part A: non-language (giving numbers to the correct pictures) • Part B: non-language (arrange the pictures by giving numbers)

	Output of Language Characteristics	—
Duration of Tasks		± 35 minutes

The objective of the third assessment was to assess the students' competence in understanding the simple story by arranging the pictures based on the story that they heard. The format of the input was given in the form of language and non-language. The story was in English as the target language and was given to the students through the audio recording. The story was relatively short and the audio speediness was slow. The input of language characteristics was restricted since the vocabulary, syntax and the function of language were limited only on the topic.

The output format was in the form of non-language because the students arranged the pictures based on the story by giving numbers for each picture. The time for doing the assessment was set into ± 25 minutes.

Table 26: The Assessment Task Characteristics of the Third Assessment

Third Assessment “Good Daily Activities”		
Assessment Objective		This designed assessment assesses students' ability to identify simple good daily activities by putting the number on the right picture based on the story in the audio recording.
Characteristics of the Input	Input of Format Characteristics	<ul style="list-style-type: none"> - Form: language (several good daily activities) and non-language (pictures) - Language: English (Target Language) - Length: Relatively short (sentences) - Type: Audio recording about good daily activities and pictorial input for interpretation - Degree of Speediness: relatively slow - Vehicle: reproduced (via audio recording)
	Input of Language Characteristics	<ol style="list-style-type: none"> 1. Language Characteristics <ul style="list-style-type: none"> - vocabulary: topical vocabulary related to the good daily activities - syntax: simple sentences - function: to identify simple daily activities 2. Topical Characteristics: restricted (good daily activities)

Characteristics of the Output	Output of Format Characteristics	- Form: non-language (arrange the pictures by giving numbers)
	Output of Language Characteristics	–
Duration of Tasks		± 25 minutes

3) Criteria of Marking

The last section was the criteria of marking. It discussed how the result of the assessment was marked. It was really needed to specify what the focus of the assessment was, how each of the focus could be described as the criteria for the test-takers and how it would be marked.

The criteria of marking consisted of the focus of competence that would be marked. The competence was described on what aspects the students had to be able to do. Each of the descriptions influenced the test specific criteria in what ways and criteria, the students would be marked.

The criteria of marking in the first assessment were focused on the language competence which was vocabulary and syntax. In terms of the vocabulary, the students were expected to understand the specific vocabulary related to the classroom instructions. The students were also asked to understand the simple imperative statements in the form of classroom instructions related to syntax. The test-takers were given 1 for each correct item and 0 for each wrong item.

Table 27: The Criteria of Marking of the First Assessment

Focus		Description	Test Specific Criteria	
Language Competence	Vocabulary	Understanding of specific vocabulary related to the classroom instructions	Test-takers are able to label each of the picture correctly based on the audio recording.	Test-takers label each of the picture which is unsuitable with the audio recording
	Syntax	Understanding simple imperative statements in the form of instructions		
Levels of Ability			1 (complete)	0 (none)

The second assessment has two criteria of marking for part A and part B. The criteria of marking for both parts were focused on the language competence. Part A concerned on vocabulary and part B concerned on vocabulary and syntax. However, each part had different correctness descriptions, different test specific criteria and also different levels of abilities.

Part A focused on the students' understanding of specific vocabulary related to the materials and the ingredients to make a cup of tea. Two points were given for each correct answer and none for the wrong answer. Part B concerned on the vocabulary related to the steps of making a cup of tea and the students' competence in understanding simple imperative statements in the form instructions. One point was for each of the correct answer and none for each of wrong answer.

Table 28: The Criteria of Marking of the Second Assessment

PART A				
Focus		Description	Test Specific Criteria	
Language Competence	Vocabulary	Understanding of specific vocabulary related to the materials and ingredients needed to make a cup of milk tea	Test-takers are able to give number on the picture correctly based on the materials and ingredients of making a cup of milk tea in audio recording	Test-takers give number to each of the picture which is unsuitable with the materials and ingredients of making a cup of milk tea in audio recording
Levels of Ability			2 (complete)	0 (none)
PART B				
Focus		Description	Test Specific Criteria	
Language Competence	Vocabulary	Understanding of specific vocabulary related to the steps of making a cup of milk tea	Test-takers are able to rearrange the pictures by giving numbers below each of the picture correctly based on the steps of making a cup of	Test-takers rearrange the pictures by giving numbers below each of the picture which is unsuitable with the steps of
	Syntax	Understanding simple imperative statements in the		

		form of instructions	milk tea in audio recording.	making a cup of milk tea in audio recording.
Levels of Ability			1 (complete)	0 (none)

The criteria of the marking for the third assessment were focused on vocabulary and syntax. Students were assessed their understanding of specific vocabulary related to the daily activities and their understanding of the simple statements in the story. One point would be given to the correct answer and none for the wrong answer.

Table 29: The Criteria of Marking of the Third Assessment

Focus		Description	Test Specific Criteria	
Language Competence	Vocabulary	Understanding of specific vocabulary related to the daily healthy habits	Test-takers are able to put the number on the right picture based on the story of daily habits correctly based on the audio recording.	Test-takers put the number on the picture based on the story of daily habits which is unsuitable with the audio recording.
	Syntax	Understanding simple statements about good daily activities		
Levels of Ability			1 (complete)	0 (none)

c. Specifications of the Listening Assessment

The listening assessment was developed based on the result of students' need analysis, the blueprint, standard of graduate competence, standard of competence and basic competence in National Curriculum. There were three topics which had different specifications. The specifications for each topic can be seen below.

a. First Assessment (*In the Classroom*)

The first assessment was "*In the Classroom*". The topic was obtained based on the material that wanted to be assessed which was classroom instructions. It was suitable with the standard of competence for the fifth grade students of

elementary school. The assessment was designed to assess the listening skill so both the worksheet and the audio recording were equally important.

Because the assessment was designed to assess listening skill, the audio recording was essential. The audio recording was used to increase validity and reliability of the assessment rather than the teacher's talk. The layout of the designed assessment was also designed by considering children characteristics and needs. It could be seen through the layout which was interesting with the colourful pictures. The instruction also used simple language so the students were able to understand easier.

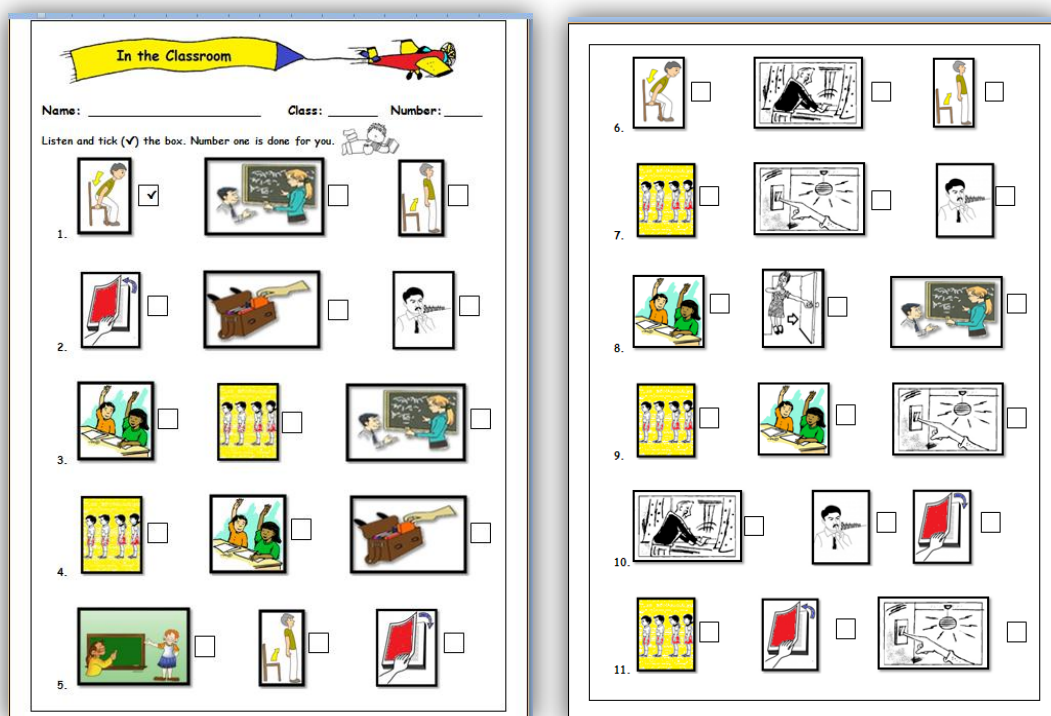


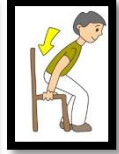


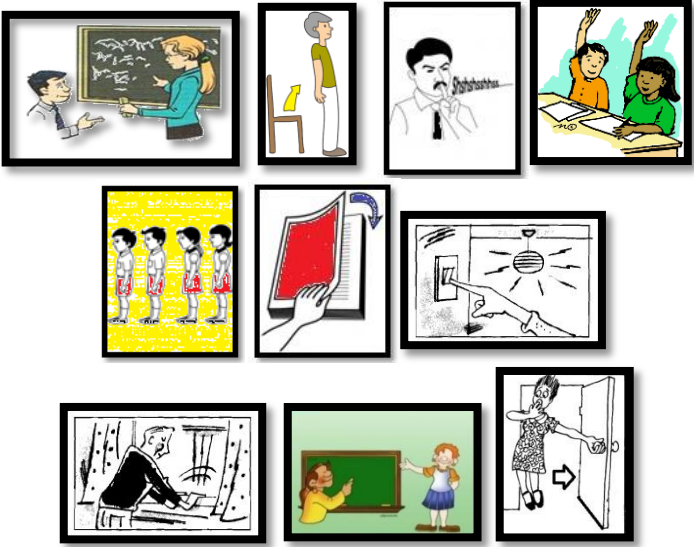
Figure 9: *In the Classroom* Assessment Design

The assessment was categorized as formal assessment since it was specifically designed, planned and organized systematically. It was also kind of the formative assessment which was aimed at evaluating the students in their learning process.

The details of the assessment specifications of the first topic can be seen in the following table.

Table 30: The Assessment Specifications of the First Assessment

Standard of Graduate Competence	Memahami instruksi, informasi dancerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar												
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah												
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah												
Skill	Listening - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macroskill</i>) - Retain chunks of language of different lengths in short-term memory (<i>microskill</i>)												
Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.												
Assessment Objective(s)	Assessing students' ability to identify the simple classroom instructions by labelling the correct pictures based on the audio recording about the classroom instructions.												
Text Types	The topic was <i>In the Classroom</i> . The assessment materials weresimple instructions which were usually done in the classroom and it would be presented in recorded form.												
Main Sub-Section of the Text	The students were given \pm 30 minutes to finish one part of the designed assessment which consisted of ten items.												
Language Competence	- vocabulary: topical vocabulary related to the classroom instructions - syntax: simple sentences and imperatives												
The Input	<p>- Language (Audio Recording)</p> <table border="0"> <tr> <td>Sit Down. (example)</td><td>Make a line.</td></tr> <tr> <td>Be Quiet.</td><td>Open the door.</td></tr> <tr> <td>Clean the board</td><td>Turn on the lamp.</td></tr> <tr> <td>Raise your hand.</td><td>Close the window.</td></tr> <tr> <td>Close your book.</td><td>Open the book</td></tr> <tr> <td>Stand up.</td><td></td></tr> </table> <p>- Non-Language (Pictures)</p> <div style="display: flex; justify-content: space-around; align-items: center;">    </div>	Sit Down. (example)	Make a line.	Be Quiet.	Open the door.	Clean the board	Turn on the lamp.	Raise your hand.	Close the window.	Close your book.	Open the book	Stand up.	
Sit Down. (example)	Make a line.												
Be Quiet.	Open the door.												
Clean the board	Turn on the lamp.												
Raise your hand.	Close the window.												
Close your book.	Open the book												
Stand up.													

	
The Expected Responses	non-language (labelling correct picture by giving a tick (√))
Criteria of Marking	<ul style="list-style-type: none"> - Each of the correct answer got 1 point - Each of the wrong answer got 0

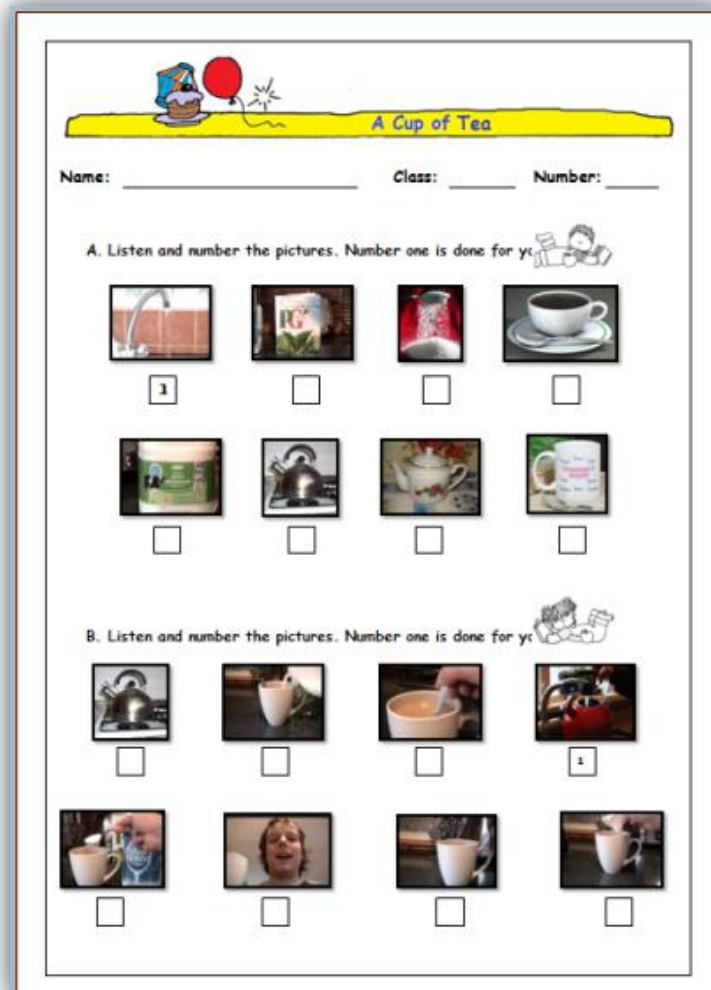
b. Second Assessment (*A Cup of Tea*)

The topic of the second assessment was “*A Cup of tea*”. It was chosen considering the standard of competence and basic competence in National Curriculum which was instructions. “*A Cup of Tea*” assessment test had simple instructions which were familiar for the students as the test-takers.

The second assessment was more complex than the first one. There were two parts that should be done by the students which had different objectives. First, the students needed to know the materials which were used to make a cup of tea. Second, they should be able to identify the steps in making a cup of tea. Because of the complexity, they were given 35 minutes to finish the designed assessment.









Same as the first one, the second assessment was also designed to assess listening skill so the audio recording was made to accompany the worksheet. The layout of the designed assessment was also colourful to create cheerful

atmosphere on the assessment process. The design can be seen in the following figure.



The worksheet is titled "A Cup of Tea" and features a yellow banner at the top with a cartoon character holding a red balloon. Below the banner, there are fields for "Name:", "Class:", and "Number:". The worksheet is divided into two sections, A and B, each with a set of eight small images and a corresponding number box.

Section A: Listen and number the pictures. Number one is done for you.

			
1			
			

Section B: Listen and number the pictures. Number one is done for you.











			
			1
			

Figure 10: A Cup of Tea Assessment Design

The second assessment was categorized as the formal assessment since the students were aware that the activities that they were doing was for the assessment purposes. It was also aimed to assess students' competencies in the process of they learn to help them in the future. The details of the assessment specifications of the second topic can be seen as follows:

Table 31: The Assessment Specification of the Second Assessment

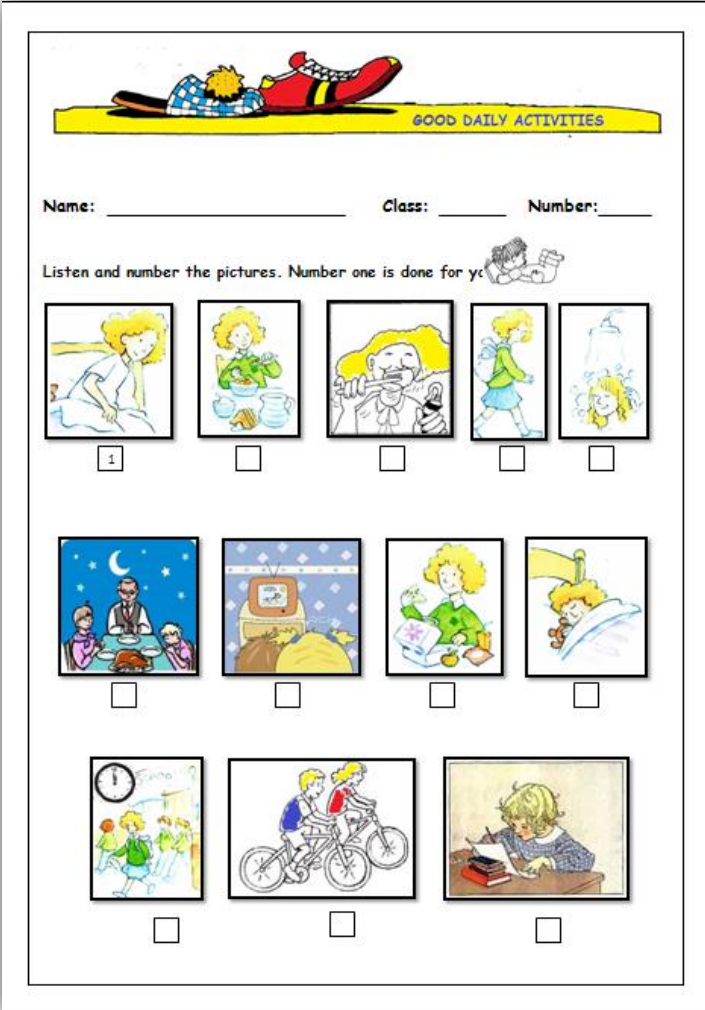
Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah
Skill	Listening - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macro skill</i>) - Retain chunks of language of different lengths in short-term memory (<i>micro skill</i>)
Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.
Assessment Objective(s)	1. Assessing students' ability to number the right pictures based on the materials and ingredients which are needed to make a cup of tea in audio recording. 2. Assessing students' ability to identify simple imperative statements by arranging and giving numbers to the pictures based on the steps of making a cup of tea in the audio recording.
Text Types	The topic was <i>A Cup of Tea</i> . The assessment materials were materials and simple instructions which were related to the ways of making a cup of tea. It would be presented in recorded form.
Main Sub-Section of the Text	The students were given \pm 35 minutes to finish two parts of the designed assessment. Part A had five items while part B had seven items.
Language Competence	- vocabulary: topical vocabulary related to the materials and steps to make a cup of tea - syntax: simple sentences and imperatives
The Input	- Language (Audio Recording) Part A some water (example) some milk a cup and a spoon some sugar a teabag a kettle Part B 1. First, put some water in the kettle. (example) 2. Second, boil the water. 3. Third, add some sugar in the cup. 4. Fourth, put a tea bag in the cup. 5. Fifth, pour the hot water into the cup carefully. 6. Sixth, add some milk in the cup. 7. Seventh, stir it firmly. 8. Finally, we get a cup of delicious tea.

	<p>- Non-Language (Pictures)</p> <p>Part A</p>  <p>Part B</p> 
The Expected Responses	<p>Part A: non-language (giving numbers to the correct pictures)</p> <p>Part B: non-language (arrange the pictures by giving numbers)</p>
Criteria of Marking	<p>Part A</p> <ul style="list-style-type: none"> - Each of the correct answer got 2 points - Each of the wrong answer got 0 <p>Part B</p> <ul style="list-style-type: none"> - Each of the correct answer got 1 point - Each of the wrong answer got 0

c. Third Assessment (*Good Daily Activities*)

The last topic of the assessment was “*Good Daily Activities*”. The topic was chosen based on the graduate competence of elementary school in National Curriculum. The third designed assessment was designed in simple language daily life which was familiar for the students.

It only had one part like the first assessment. The materials were presented in simple story which was related to the daily life. It was suitable with the assessment objective which assessing students’ competencies by numbering the picture based on the story of daily activities they heard.



GOOD DAILY ACTIVITIES

Name: _____ Class: _____ Number: _____

Listen and number the pictures. Number one is done for you.











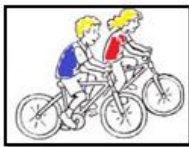


 1				
				
				

Figure 11: Good Daily Activities Assessment Design

The last assessment was also designed to assess listening skill. The activity belonged to the selective listening which means that the students processed a short monologue in order to get certain information that they need. The details of the assessment specifications of the third topic can be seen as follows:

Table 32: The Assessment Specifications of the Third Assessment

Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah

Skill	<p>Listening</p> <ul style="list-style-type: none"> - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macroskill</i>) - Retain chunks of language of different lengths in short-term memory (<i>microskill</i>)
Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.
Assessment Objective(s)	Assessing students' ability to identify simple good daily activities by putting the number on the right picture based on the story in the audio recording.
Text Types	The topic was <i>Good Daily Activities</i> . The assessment materials were several daily activities which were usually done and it would be presented in recorded form.
Main Sub-Section of the Text	The students were given \pm 25 minutes to finish one part of the designed assessment which consisted of eleven items.
Language Competence	<ul style="list-style-type: none"> - vocabulary: topical vocabulary related to good daily activities - syntax: simple sentences
The Input	<p>- Language (Audio Recording)</p> <p>I get up at five o'clock in the morning. Then, I take a bath and get dressed. I have breakfast before I go to school. I brush my teeth after having it. Then, I go to school. At the break time, I have my lunch. I go home at twelve o'clock. I have a break then I ride my bicycle with my brother. After that, I do my homework. At seven o'clock, I have dinner. I enjoy my evening by watching TV with my family. At nine o'clock, I go to bed.</p> <p>- Non-Language (Pictures)</p> 
The Expected Responses	non-language (arrange the pictures by giving numbers)
Criteria of Marking	<ul style="list-style-type: none"> - Each of the correct answer got 1 point - Each of the wrong answer got 0

2. Getting Expert Judgement, Evaluating and Revising the First Draft of the Listening Assessment

Three topics of the listening assessment which were as the first draft had been designed based on the blueprint, the result of students' need analysis, the standard of graduate competence, standard of competence and the basic competence in National Curriculum. Before it was tried-out, it is needed to get suggestions and improvements from the experts.

a. Getting Expert Judgement of the First Draft of the Listening Assessment

Suggestions and improvements from the experts were got through the evaluation questionnaire. The experts were given an evaluation questionnaire for each topic which consisted of 20 items. The blueprint of the evaluation questionnaire can be seen below.

Table 33: Blueprint of the Evaluation Questionnaire Administered to Respondents

The Purpose of the Questions	Part	Question Number	References
To find some information related to the respondents' profile	A	-	Tomlinson (1998: 240)
To find some information to evaluate the correlation between the designed assessment and the National Curriculum	B	1, 2	Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar (2008: 11, 17-19)
To find some information to evaluate validity of the designed assessment		3, 4, 5, 6, 7, 8, 9, 10	Bachman & Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation & Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment		11, 12, 13, 14, 17	Gordon (2007: 202); Nation & Newton (2009: 167); Brown (2004: 31-32)
To find some information to evaluate practicality of the designed assessment		15	Bachman & Palmer (1996: 155); Baxter (1997: 28); Nation & Newton (2009: 168); Brown (2004: 31)
To find some information to evaluate the authenticity and the interactiveness of the designed assessment		16	Bachman & Palmer (1996: 151-152); Brown (2004: 35)
To find some information to		18, 20	Bachman & Palmer (1996: 153-

evaluate backwash effect of the designed assessment			155); Baxter (1997: 28-29)
To find some information to evaluate the fairness of the designed assessment		19	Cameron (2001: 226); Wakeford (2003: 45-46)

b. Evaluating the First Draft of the Listening Assessment

The first part of the questionnaire obtained the information related to the respondents' profile. There were two experts as the respondents. Both of them have lectured in English Education Department of Yogyakarta State University for more than five years.

The second part consisted of seven elements which collected the data related to the evaluation of the first draft. They were the correlation between the designed assessment and the National Curriculum, the validity, the reliability, the practicality, the authenticity and interactiveness, the backwash effect and the fairness of the designed assessment. The questionnaire fulfilled the content validity since the items were written in line with the elements which had already designed in the blueprint.

1) First Assessment (*In the Classroom*)

The first topic of the assessment was "*In the Classroom*". It was evaluated by the experts through the seven elements of the questionnaire. It regarded reliable since the Alpha's score was 0.877. The data computation details can be checked in Appendices while the further description about the result of the questionnaire can be seen below.

The first element was the correlation between the designed assessment and the National Curriculum. According to the experts, the first topic was already suitable with the National Curriculum in terms of the standard of graduate competence of

elementary school students, the standard of competence and the basic competence of fifth grade students of elementary school. Materials, which were assessed, were simple instructions as what have been mentioned in National Curriculum.

Table 34: Experts' Responses on Evaluation Questionnaires of the First Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	1	3.50	Very Good
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	2	3.50	Very Good

The second element was related to the validity of the designed assessment. According to the experts, the first topic had already fulfilled the validity criteria. The items in the designed assessment were rewritten in line with the blueprint of the assessment. The contents were also suitable with the students' needs, their language competencies, their characteristics, and culture since the topic was "*In the Classroom*" which was familiar for them. Moreover, the first topic of the designed assessment was able to measure of what really wanted to be measured which was listening skill. In terms of the layout, the pictures and the font size could be seen clear for the students because both of them were big enough to see. The design was also interesting with coloured pictures which were colourful.

Table 35: Experts' Responses on Evaluation Questionnaires of the Second Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment was suitable with the blueprint designed.	3	3.50	Very Good
	The designed assessment was suitable with the students' needs.	4	3.00	Good
	The designed assessment was suitable with the students' language competencies.	5	3.00	Good
	The designed assessment was suitable with the students' background, their ages, the culture and their hobbies.	6	3.00	Good
	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	7	3.50	Very Good
	The designed assessment had the interesting layout.	8	3.00	Good
	The pictures in the designed assessment could be seen clearly.	9	3.00	Good
	The font size in the designed assessment was readable.	10	3.00	Good

The third element was the reliability of the designed assessment. In this case, the number of items in first topic of the designed assessment needed to be revised. The experts gave suggestions to add the items in the worksheet since there were only four items. It would be better when number of items for all the designed assessment was balanced. The instructions were clear for the students since it was simple and only consisted of simple sentences with familiar vocabulary. The scoring rubric was also provided for the first topic of the designed assessment. The criteria of marking in the scoring rubric were written in line with the aspects that wanted to be assessed in the first topic. The audio recording was also audible for the students as the test-takers.

Table 36: Experts' Responses on Evaluation Questionnaires of the Third Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	11	2.50	Fair
	The designed assessment had clear instructions.	12	3.50	Very Good
	There was the scoring rubric for the designed assessment.	13	3.00	Good
	The audio recording in the designed assessment was audible.	14	3.00	Good
	The criteria of marking were suitable with the designed assessment.	17	3.00	Good

The fourth element of the questionnaire was the practicality of the designed assessment. In this aspect, the experts agreed that the time allocation for the first topic of the assessment was enough. The students did not need too much time since the difficulty level of the first assessment was relatively easy. Therefore, it was visible to do.

Table 37: Experts' Responses on Evaluation Questionnaires of the Fourth Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the practicality of the designed assessment	Time allocation for the designed assessment was enough.	15	3.00	Good

The next important element was the authenticity and the interactiveness of the designed assessment. According to the experts, the first topic of the designed assessment already fulfilled the element. The teacher often introduced them with kinds of classroom instructions. Therefore, it was suitable with the students' daily condition.

Table 38: Experts' Responses on Evaluation Questionnaires of the Fifth Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the authenticity and the interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	16	3.00	Good

The sixth element was evaluating the backwash effect in the designed assessment. Both of the experts agreed that the first topic of the designed assessment was able to motivate the students more to learn English. It was also able to enhance the students' competences in their listening skill.

Table 39: Experts' Responses on Evaluation Questionnaires of the Sixth Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the backwash effect of the designed assessment	The designed assessment was able to motivate students to learn English.	18	3.00	Good
	The designed assessment was able to help the students to enhance their competencies in listening skill.	20	3.00	Good

The last was the fairness of the designed assessment. In terms of the audio recording, the first topic of the designed assessment already fulfilled the aspect of fairness because the audio was clear for the students. Moreover, the worksheet which was given was comfortable to read for all of the students.

Table 40: Experts' Responses on Evaluation Questionnaires of the Seventh Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the fairness of the designed assessment	The designed assessment fulfilled the aspect of fairness for the students.	19	3.00	Good

2) Second Assessment (*A Cup of Tea*)

The second topic of the designed assessment was “*A Cup of Tea*”. The topic was also evaluated by the evaluation questionnaire. It was regarded reliable since the Alpha’s score was 0.921. Further descriptions related to the result of the questionnaires can be seen below.

The first element was the correlation between the designed assessment and the National Curriculum. According to the experts, the second topic of the designed listening assessment was already suitable with the standard of graduate competence, standard of competence and basic competence for the elementary school students. Materials that wanted to be assessed were simple instructions in the form of simple procedure texts. The detail of data computation can be seen in the following table.

Table 41: Experts’ Responses on Evaluation Questionnaires of the First Element “*A Cup of Tea*” by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	1	3.50	Very Good
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	2	3.50	Very Good

The second element was the validity of the designed listening assessment. According to the experts, the designed assessment had already suitable with the blueprint that was designed. It also fulfilled the students’ need and was able to assess the students’ language competencies. The items also assessed what should

be assessed and had been already arranged into an interesting designed listening assessment by providing colourful pictures with readable font size.

Table 42: Experts' Responses on Evaluation Questionnaires of the Second Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment was suitable with the blueprint designed.	3	3.50	Very Good
	The designed assessment was suitable with the students' needs.	4	3.00	Good
	The designed assessment was suitable with the students' language competencies.	5	3.00	Good
	The designed assessment was suitable with the students' background, their ages, the culture, and their hobbies.	6	3.50	Very Good
	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	7	3.00	Good
	The designed assessment had the interesting layout.	8	3.00	Good
	The pictures in the designed assessment could be seen clearly.	9	3.00	Good
	The font size in the designed assessment was readable.	10	3.00	Good

The third element was the reliability of the assessment. The second designed listening assessment was regarded reliable since it fulfilled the characteristics of the assessment for young learners. The number of items was appropriate with the age of the students because the items were not too many. The instructions were also simple and cleared enough for the test-takers. It was consisted of simple sentences and vocabularies which accompanied with the symbol on how giving a tick to make it easier for the students. The second assessment had the scoring rubrics and the criteria of the correctness was clear, easy to understand and

suitable with the aspects that wanted to be assessed. Moreover, the audio recording was audible for the students.

Table 43: Experts' Responses on Evaluation Questionnaires of the Third Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	11	3.00	Good
	The designed assessment had clear instructions.	12	3.50	Very Good
	There was the scoring rubric for the designed assessment.	13	3.00	Good
	The audio recording in the designed assessment was audible.	14	3.00	Good
	The criteria of marking were suitable with the designed assessment.	17	3.50	Very Good

The fourth element was the practicality of the designed listening assessment. In this case, the experts agreed that the second topic was feasible to do. Time allocation which was given to the students was enough for them. They were given ± 30 minutes because it was considered having quite many items.

Table 44: Experts' Responses on Evaluation Questionnaires of the Fourth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the practicality of the designed assessment	Time allocation for the designed assessment was enough.	15	3.00	Good

The fifth element in the evaluation questionnaire for the experts was the authenticity and the interactiveness of the designed listening assessment. To fulfil those elements, the designed listening assessment should be suitable with the students' daily life. It means that they could find it in the real life. Ingredients and steps of making a cup of tea were considered authentic for the students.

Table 45: Experts' Responses on Evaluation Questionnaires of the Fifth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the authenticity and the interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	16	3.00	Good

The sixth element was the backwash effect of the designed listening assessment. According to the experts, the second assessment gave good backwash effect for the students. It was able to motivate the students to learn English more and enhanced their listening skill competencies. It could be seen through the characteristics of the worksheet which was fun and colourful for the students.

Table 46: Experts' Responses on Evaluation Questionnaires of the Sixth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the backwash effect of the designed assessment	The designed assessment was able to motivate students to learn English.	18	3.00	Good
	The designed assessment was able to help the students to enhance their competencies in listening skill.	20	3.00	Good

The last important element was the aspect of fairness in the designed listening assessment. All of the aspects of the second designed listening assessment already fulfilled the criteria of fairness since the worksheet was able to promote the same chance for the students in the assessment process. The second topic was also possible to do in a big class and give the same chance for every student.

Table 47: Experts' Responses on Evaluation Questionnaires of the Seventh Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the fairness of the designed assessment	The designed assessment fulfilled the aspect of fairness for the students.	19	3.00	Good

3) Third Assessment (*Good Daily Activities*)

The third assessment was "*Good Daily Activities*". It was approved and had been evaluated by the experts like the first and second assessment through the evaluation questionnaire. The questionnaire was reliable since the Alpha's score was 0.902.

The first element of the questionnaire was the correlation between the designed listening assessments and the National Curriculum. According to the experts, the items and the content on third assessment were already appropriate with the standard of graduate competence, standard of competence and the basic competence in the National Curriculum. It could be checked through the standard of graduate competence for elementary school students which have been mentioned that understanding the simple story was one of the content.

Table 48: Experts' Responses on Evaluation Questionnaires of the First Element "*Good Daily Activities*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	1	3.50	Very Good
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	2	3.50	Very Good

The second element was the validity of the designed assessment. The result of the questionnaire showed that the third topic of the designed assessment was valid. It could be seen through the items that were suitable with the blueprint. All of the items also were designed based on the students' characteristics and needs. It had the interesting layout and colourful pictures that could motivate the students to learn English.

Table 49: Experts' Responses on Evaluation Questionnaires of the Second Element "*Good Daily Activities*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment was suitable with the blueprint designed.	3	3.50	Very Good
	The designed assessment was suitable with the students' needs.	4	3.50	Very Good
	The designed assessment was suitable with the students' language competencies.	5	3.00	Good
	The designed assessment was suitable with the students' background, their ages, the culture and their hobbies.	6	3.00	Good
	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	7	3.00	Good
	The designed assessment had the interesting layout.	8	3.00	Good
	The pictures in the designed assessment could be seen clearly.	9	3.00	Good
	The font size in the designed assessment was readable.	10	3.00	Good

The third element in the evaluation questionnaire was the reliability of the designed listening assessment. The experts had evaluated that the number of items in the third assessment was suitable for young learners. The instructions were also clear since the instructions were accompanied by a clear example like the first and second questionnaire. Moreover, it accompanied by the appropriate scoring rubric

with specific criteria of the correctness. The audio recording was also audible for the students.

Table 50: Experts' Responses on Evaluation Questionnaires of the Third Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	11	3.00	Good
	The designed assessment had clear instructions.	12	3.50	Very Good
	There was the scoring rubric for the designed assessment.	13	3.00	Good
	The audio recording in the designed assessment was audible.	14	3.00	Good
	The criteria of marking were suitable with the designed assessment.	17	3.50	Very Good

The fourth element was the practicality of the designed listening assessment. It was related to the fact whether the third topic of the assessment was feasible to do or not. According to the experts, the time allotment was enough for the students considering the level of difficulty and number of items.

Table 51: Experts' Responses on Evaluation Questionnaires of the Fourth Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the practicality of the designed assessment	Time allocation for the designed assessment was enough.	15	3.00	Good

The fifth element was the authenticity and the interactiveness of the designed listening assessment. The experts agreed that the third topic of the designed listening assessment was suitable with the students' daily life because the materials were the common activities that usually done by the students.

Table 52: Experts' Responses on Evaluation Questionnaires of the Fifth Element "*Good Daily Activities*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the authenticity and the interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	16	3.00	Good

The sixth element was the backwash effect of the designed listening assessment. According to the experts, the third assessment was able to motivate the students to learn English because the items were suitable with the students' characteristics. It also helped the students to increase their competencies in listening skill.

Table 53: Experts' Responses on Evaluation Questionnaires of the Sixth Element "*Good Daily Activities*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the backwash effect of the designed assessment	The designed assessment was able to motivate students to learn English.	18	3.50	Very Good
	The designed assessment was able to help the students to enhance their competencies in listening skill.	20	3.00	Good

The last was the fairness of the designed assessment. In terms of the audio recording, the first topic of the designed assessment already fulfilled the aspect of fairness because the audio was clear for the students. Moreover, the worksheet which was given was comfortable to read for all of the students.

Table 54: Experts' Responses on Evaluation Questionnaires of the Seventh Element "*Good Daily Activities*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the fairness of the designed assessment	The designed assessment fulfilled the aspect of fairness for the students.	19	3.00	Good

In the validation sheet, the experts also add some suggestions for all of the designed listening assessment topics. First, it would be better if the sounds in audio recording were the children's voice to increase the cheerful atmosphere of the assessment. Another reason was that children loved to have Relationship with their pairs. Second, the time interval between the sentences in audio recording should be given more to give enough time for the students to understand and think about the answers.

c. The Revision of the First Draft of the Listening Assessment

The result of the evaluation questionnaire for the experts and suggestions from them were used as the basic to revise the designed listening assessment. The revision can be seen in the following table.

Table 55: The General Revision Table of the First Draft of the Designed Listening Assessment

	Feedback	Improvement
First Assessment (In the Classroom)	There should be more items in the worksheet.	The items in the worksheet had been already added more
	It would be better to use children's voice.	The audio recording still used adult's voice because of the technical skill. However, the pitch and tone were adjusted to make the voice more familiar for the students.

	The time interval in audio recording should be added more.	The time interval was already added more.
Second Assessment (A Cup of Tea)	It would be better to use children's voice.	The audio recording still used adult's voice because of the technical skill. However, the pitch and tone were adjusted to make the voice more familiar for the students.
	The time interval in audio recording should be added more.	The time interval was already added more.
Third Assessment (Good Daily Activities)	It would be better to use children's voice.	The audio recording still used adult's voice because of the technical skill. However, the pitch and tone were adjusted to make the voice more familiar for the students.
	The time interval in audio recording should be added more.	The time interval was already added more.

3. Trying Out, Evaluating and Revising the Second Draft of the Listening Assessment

After getting expert judgements, the first draft of the designed listening assessment was revised in some parts. The result was the second draft which should be tried-out. The try-out was conducted on September 2013 covering the three topics of the assessment. Each topic was tried-out twice. The data from the evaluation questionnaire, observation sheets and the interviews to revise the second draft were gained in the first try-out together with the assessment scores. In the second try-out, the data of the assessment scores were collected once more to analyze the reliability of the designed assessment.

To evaluate the second draft, the data from evaluation questionnaire for students were used. The blueprint of the evaluation questionnaire can be seen in the following table.

Table 56: Blueprint of the Evaluation Questionnaire administered to Students

The Purpose of the Questions	Part	Question Number	References
To find some information related to students' profile	A		Tomlinson (1998: 240)
To get some information related to the students attitude toward the designed assessment	B	2, 14	Georgiou & Pavlou (2003: 4-6, 8)
To find some information related to the validity of the designed assessment.		1, 4, 5, 6	Bachman & Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation & Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment		3, 7, 8, 9	Gordon (2007: 202); Nation & Newton (2009: 167); Brown (2004: 31-32)
To find some information related to the practicality of the designed assessment		10, 11	Bachman & Palmer (1996: 155); Baxter (1997: 28); Nation & Newton (2009: 168); Brown (2004: 31)
To find some information related to the authenticity and the interactiveness of the designed assessment		12	Bachman & Palmer (1996: 151-152); Brown (2004: 35)
To find some information about the backwash effect of the designed assessment		13, 15	Bachman & Palmer (1996: 153-155); Baxter (1997: 28-29)

The data from the observation sheet were gained from the teacher and the observer. There was also the interview session with the teacher and students after the first try-out of each assessment topic. The blueprint of the observation sheet can be seen in the following table.

Table 57: Blueprint of the Classroom Observation Sheet

The Purpose of the Questions	Question Number	References
To find some information related to respondents' profile		Tomlinson (1998: 240)
To get some information related to the students attitude toward the designed assessment	1, 3	Georgiou & Pavlou (2003: 4-6, 8)
To find some information about the correlation between the designed assessment and the national curriculum	5, 6	Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar (2008: 11, 17-19)

To find some information to evaluate validity of the designed assessment	7, 8, 9, 10, 11, 12, 13	Bachman& Palmer (1996: 150-151); Cameron (2001: 225); Baxter (1997: 18-22); Nation& Newton (2009: 167-168)
To find some information to evaluate reliability of the designed assessment	14, 15	Gordon (2007: 202); Nation& Newton (2009: 167); Brown (2004: 31-32)
To find some information about practicality of the designed assessment	16	Bachman& Palmer (1996: 155); Baxter (1997: 28); Nation& Newton (2009: 168); Brown (2004: 31)
To find some information about the authenticity and the interactiveness of the designed assessment	17	Bachman& Palmer (1996: 151-152); Brown (2004: 35)
To find some information about the backwash effect of the designed assessment	18, 19	Bachman& Palmer (1996: 153-155); Baxter (1997: 28-29)
To find some information about the fairness of the designed assessment	2, 4	Cameron (2001: 226); Wakeford (2003: 45-46)

There were some aspects which were evaluated based on the try-out. The first aspect is the students' attitude toward the designed assessment. Second, it is related to the relevance between the designed listening assessment and the National Curriculum. The last aspect is the principles of the appropriate assessment covering validity, reliability, practicality, the authenticity, interactiveness, backwash effect and the fairness.

a. Trying Out, Evaluating and Revising the Second Draft of the First Assessment Topic

1) The Try-Out

The first assessment topic was *In the Classroom*. It was tried out twice on September 16 and 30, 2013. The first try-out was followed by all of the students while the second try-out was followed by 37 students.

2) The Evaluation

a) Students' Attitude toward the Second Draft Assessment Designed

The data of students' attitude toward the designed assessment were obtained through the evaluation questionnaire. The result can be seen in the following table.

Table 58: Students' Responses on Evaluation Questionnaires of the First Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
Students' attitude toward the designed assessment	I loved the English assessment which was given.	2	2.98	Good
	I was not afraid anymore to finish the English assessment.	14	3.07	Good

As seen in the table 58, the mean score of question number two was 2.98 and the mean score of question number fourteen was 3.07. Those were categorized as good. It means that the students had a good attitude toward the designed assessment. Another supporting data came from the excerpt below.

.....
 R : ..., *mohon untuk bisa diberikan tanggapan bu untuk tes hari ini.* (..., what is your opinion about the assessment today?)

T : ..., *Kalau menurut saya, anak-anak sih suka mbak soalnya beda.* (..., in my opinion, the students love the assessment because it is different.)

(Appendix E: The Interview Transcripts of Assessment I)

.....
 The data of the interview transcript above was between the teacher and I. She concluded that the students loved the assessment. The reason was because the designed assessment was different from the assessment they usually faced. She also stated the similar opinion in the observation sheet. The result is presented below.

Table 59: The Teacher's Responses on Observation Sheet of the First Element "In the Classroom" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
Students' attitude toward the designed assessment	Students were comfortable when doing the designed assessment.	1	Some of them were afraid but after they saw the designed assessment, they looked happy.
	Students were having more spirits to face the designed assessment.	3	Most of them were enthusiastic although some of them were not.

The data about the students' attitude towards the designed assessment were also gained from the interview with the students. One of the students said that she liked the designed assessment. The students love the assessment because the layout of the worksheet is interesting for her. The data can be seen in the following excerpt.

-
- R : *Nah, kakak mau tanya nih tentang tes yang tadi kakak berikan. Gimana menurut adik?* (Well, I want to ask you about the assessment that I gave. What's your opinion?)
- S : *Wah, kalo aku sih suka kak soalnya tesnya lucu.* (I loved the assessment because it was funny.)
- R : *Lucunya yang gimana?* (What do you mean?)
- S : *Ada gambarnya gitu. Terus gak hitam putih.* (Because there were pictures and it was colourful.)

(Appendix E: The Interview Transcripts of Assessment I)

.....

The data shows that the students had a good attitude toward the first assessment topic. According to the teacher, at first, they looked afraid because they knew that they would face the designed assessment. However, after they saw the worksheet and felt the assessment atmosphere, most of them looked enthusiastic.

b) Relationship between the National Curriculum and the Second Draft Assessment Designed

The result of the data analysis shows that the designed assessment is relevance with the National Curriculum. The teacher said that the assessment tasks have been made based on the Standard Competence and Basic of Competence. The data were gained through the interview with the teacher as follows:

R :*Tentang materinya bagaimana?*(OK, What do you think about the materials?)

T : *sudah sesuai kok. Kan instruksi di kelas jadi ya sudah sama dengan yang ada di SK KD.*(It is suitable with the National Curriculum. The materials are the classroom instructions. So, it was similar to the Standard of Competence and Basic Competence.)

(Appendix E: The Interview Transcripts of Assessment I)

In the observation sheet, both of the teacher and the observer were agree that the first assessment topic was already suitable with the National Curriculum. The result can be seen in the following table.

Table 60: Respondents' Responses on Observation Sheet of the Second Element "In the Classroom" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	5	It was suitable.
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	6	It was suitable.

c) Validity

Validity of the designed assessment is measured using three types of validity which are content validity, construct validity and face validity. The data were collected from the teacher, the students and the observer. The explanation about validity is given below.

i) Content Validity

The content validity of the designed assessment is related to the aspect whether the designed assessment assesses the competencies that should be assessed or not. The data related to the content validity were obtained from the observation sheet and the evaluation questionnaire.

The items number 7, 8 and 9 in the observation sheet concerned on the content validity. The data from the teacher shows that the designed assessment fulfilled the criteria of content validity. It can be seen from her responses to the items in the table below.

Table 61: The Teacher's Responses on Observation Sheet of the Third Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was suitable with the students' needs.	7	Students were happy with the interesting one so it was suitable.
	The designed assessment was suitable with the students' background, their ages, the culture, and their hobbies.	8	It was suitable.
	The designed assessment was suitable with the students' language competencies.	9	Vocabulary which was used was simple. So, the students were easy to understand.

In the evaluation questionnaire, the item number 1 is related to the content validity. The mean score was 3.15 which fall in the good category. It means that the students agree with the statement.

Table 62: Students' Response on Evaluation Questionnaires of the Third Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment which was given was suitable with what I have learned	1	3.15	Good

ii) Construct Validity

Construct validity focused on the appropriateness of the assessment procedures with the competencies that wanted to be assessed. The item number 11 in the observation sheet asks the responses from the teacher and the observer related to the construct validity. The observer said that the assessment process showed the ways to assess listening skill. The result can be seen below.

Table 63: The Observer's Response on Observation Sheet of the Third Element "In the Classroom" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	11	The activities really showed that it was listening designed assessment.

The English teacher had the similar opinion that the designed listening assessment achieved the content validity. Her opinion was collected in the interview session after the first try-out.

R :*Selain materi, apakah prosedur pemberian tesnya sudah sesuai bu dengan apa yang diteskan?*(Besides the assessment tasks, do you think the assessment procedure is suitable with the task be assessed ?)

T : *Anak-anak itu mencentang gambar setelah mendengar instruksi. Jadi, menurut saya sudah bisa membuka kemampuan listening.* (The students ticked the pictures after heard the instructions. So, I think it could measure the students listening competencies.)

(Appendix E: The Interview Transcripts of Assessment I)

iii) Face Validity

Face validity concerned on how the assessment appears. Students evaluated the face validity aspect through the evaluation questionnaire given to them in the first try out. The mean scores of items number 4, 5 and 6, which represent the face validity, are 3.41, 3.07 and 2.98. Those fall into very good and good category. It means that the layout of the assessment fulfilled the criteria of the face validity.

Table 64: Students' Response on Evaluation Questionnaires of the Third Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	I loved the designed assessment because it was colourful.	4	3.41	Very Good
	I loved the designed assessment because there were a lot of pictures.	5	3.07	Good
	I could read the designed assessment easily.	6	2.98	Good

The English teacher stated her opinion in item number 10, 12 and 13 in the observation sheet which related to face validity. Overall, she gave good responses. However, the pictures in the worksheet need the improvement. The data are showed in the following table.

Table 65: Respondents' Responses on Observation Sheet of the Third Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment had the interesting layout.	10	The worksheet was so colourful.
	The pictures in the designed assessment could be seen clearly.	12	The pictures were too small.
	The font size in the designed assessment was readable.	13	It was clear.

d) Reliability

Reliability of the designed assessment measures how well an assessment assesses what it claims. The data were gained from the various sources.

First, the assessment scores on the first try-out were analyzed with the assessment scores on the second try-out to get the reliability coefficient of the designed assessment. The result of the reliability coefficient, which was measured using the Alpha's Cronbach statistics from the two test scores, is 0.890. It shows that the first assessment topic is reliable.

Second, the data were collected from the evaluation questionnaire for the students. The assessment reliability also depends on the physical context of the assessment. The data is presented in the following table.

Table 66: Students' Response on Evaluation Questionnaires of the Fourth Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the reliability of the designed assessment	The designed assessment was not too difficult to do.	3	3.12	Good
	The items of the designed assessment were not too many to do.	7	3.49	Very Good
	I understood the instructions of the designed assessment.	8	3.02	Good
	I could hear the sound clearly.	9	2.93	Good

The mean scores in the table show that the designed listening assessment is reliable. All of them were in the level of the good and very good. In general, the physical context of the designed assessment is good.

The other supporting data were from the result of the observation sheet and the interview session with the students. The data of the observation sheet administered to the observer shows that the items in the worksheet were enough, the audio recording was clear but the time interval should be lengthen.

Table 67: The Observer's Response on Observation Sheet of the Fourth Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	14	Yes
	The audio recording in the designed assessment was audible.	15	The audio recording was good but it should be lengthen.

Based on the data of the teacher, the audio recording was good and the items in the worksheet were also enough. However, she said that the sequence of the items should be revised.

Table 68: The Teacher's Responses on Observation Sheet of the Fourth Element "In the Classroom" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	14	The number of items was enough but the sequence of number should be revised.
	The audio recording in the designed assessment was audible.	15	Yes

One of the students said that she had the problem with the audio recording.

The audio was clear but she was unfamiliar with the native speaker. However, after I gave more explanations and helped her, she said that she understood.

R : *kan tesnya mendengarkan, nah, suaranya kedengaran jelasgak?* (The designed assessment is listening. Is the audio recording clear?)

S : *Denger sih tapi agak susah. Suaranya gimana gitu.* (It is clear but it is quite difficult.)

R : *Gimana apanya?* (What do you mean?)

S : *Bule banget. Kan jarang denger kak.* (The sounds is native speaker. I rarely heard that kind of voice.)

R : *Kalo gitu, waktu tadi kakak jelasin lagi, jadi lebih gampang gak?* (After I give more explanation, is it easier?)

S : *Iyalah. Kan jad imudeng gitu aku.* (Yes it is. I understand.)

(Appendix E: The Interview Transcripts of Assessment I)

e) Practicality

The standardized listening assessment should be feasible to do. The practicality aspect can be seen from the time which is given to the test-takers and whether the scoring system is feasible in the assessor's time frame or not.

In the aspect of the allocation time, the students gave positive responses.

Based on the data, the mean scores of the items fall into good category.

Table 69: Students' Responses on Evaluation Questionnaires of the Fifth Element "In the Classroom" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The practicality of the designed assessment	I could finish all of the items in the designed assessment.	10	3.02	Good
	I could finish all of the items in the designed assessment on time.	11	3.00	Good

The other data related to the time allocations were gained through the interview with the teacher and one of the students. The excerpts can be seen as follows:

.....
 R :*Kalau tentang waktu yang diberikan bagaimana?*(What do you think about the time allocation for the students?)

T : *Kalau bisa, waktunya lebih lama mbak. Apalagi tes seperti ini kan termasuk baru untuk dikenalkan ke anak.*(It will be better to add the time allocation because the designed assessment is new for the students.)

(Appendix E: The Interview Transcripts of Assessment I)

.....
 R : *Kalo waktu untuk mengerjakannya cukupgak?*(Do you think the allocation time for students is enough?)

S : *Cukup kok. Kan soalnya gak susah.*(Yes, it is because the items are easy.)

(Appendix E: The Interview Transcripts of Assessment I)

.....
 The data related to the assessor's time frame in scoring system were obtained from the interview session with the teacher. She said that the scoring system was simple and she could use it effectively in assessing the students' competencies in listening skill.

.....
 R : *Bagaimana menurut ibu tentang cara penilaian tes ini?* (What do you think about the scoring system of the designed assessment?)

T : *Menurut saya cukup sederhana mbak. Selain itu juga cepat dan efektif.*(I think it is quite simple. Besides, the scoring system also fast and effective.)

(Appendix E: The Interview Transcripts of Assessment I)

f) Authenticity and Interactiveness

An appropriate assessment should be relevant with the students' daily life. The assessment tasks also should be interactive which means that the designed

assessment should involve the test takers' individual characteristics in accomplishing the assessment tasks. The mean score of the item number 12 in the evaluation questionnaire for students is 3.49. It proven that the designed assessment is authentic and interactive.

Table 70: Students' Response on Evaluation Questionnaires of the Sixth Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The authenticity and interactiveness of the designed assessment	I could find the designed assessment materials in my daily life.	12	3.49	Very Good

The data above were supported by the result of classroom observation sheet which were administered to the teacher and the observer as the respondents. Both of them gave similar opinion. The data are shown in the following tables.

Table 71: The Teacher's Responses on Observation Sheet of the Sixth Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
The authenticity and interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	17	It was suitable so the students could understand easier.

Table 72: The Observer's Responses on Observation Sheet of the Sixth Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
The authenticity and interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	17	Yes

g) Backwash Effect

In designing the appropriate assessment, the backwash effect should be considered. It should achieve good backwash effect rather than the bad one. The

observer suggested the feedback should be given in the various ways. The designed assessment also could help the students to improve their listening competencies.

Table 73: The Observer's Responses on Observation Sheet of the Seventh Element "*In the Classroom*" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
The backwash effect of the designed assessment	The designed assessment could help the students to develop their abilities in listening skill.	18	The designed assessment could help the students.
	The feedback which was given was able to improve students' competencies.	19	It would be better if the feedback was more various.

The data related to the backwash effect of the designed assessment were also gained from the students. The mean scores of the items number 13 and 15 which represented the backwash effect were 2.98 and 2.93. Those fall into good category. Based on the data on table 74, the students agreed that the designed assessment gave good backwash effect for them.

Table 74: Students' Responses on Evaluation Questionnaires of the Seventh Element "*In the Classroom*" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The backwash effect of the designed assessment	I loved English more.	13	2.98	Good
	The designed assessment helped me to learn more about English.	15	2.93	Good

h) Fairness

All of the aspects above would not effective unless all of the students had the same chances to do the designed assessment. The teacher agreed that the designed assessment fulfilled the aspect of fairness. The data can be seen as follows:

R : Menurut ibu, apakah semua siswa di kelas sudah mendapatkan kesempatan yang sama bu dalam tes ini?(Do you think all of the students in the class get the same chances to do the designed assessment?)

T : Kalau menurut saya sih sudah mbak karena lembar tes semua dapat, yang dibelakang juga bisa dengar suaranya dengan jelas.(In my opinion, they get the same chances because all of them get the worksheet and the students who sit behind can hear the audio recording clearly.)

(Appendix E: The Interview Transcripts of Assessment I)

The supporting data was also got through the observation sheet. The result in the following table was obtained from the English teacher as the respondent.

Table 75: Respondents' Responses on Observation Sheet of the Eighth Element "In the Classroom" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
Fairness of the designed assessment	Every student got the same facilities (worksheet, audio recording, etc) in doing the designed assessment.	2	Yes
	Every student had the same chances to do the designed assessment.	4	Yes

3) The Revision

Based on the analysis of the try-out of the first assessment topic, generally the designed assessment already fulfilled the criteria for all of the aspects. However, there were some parts which needed to be improved. The revisions can be seen in the following table.

Table 76: The Revision of the Assessment Aspects of the First Assessment Topic

Aspects that need to be revised	Revisions
Students' Attitude toward the Designed Assessment	-
Relationship between the National Curriculum with the Designed Assessment	-
Validity	Make the pictures in the worksheet bigger.
Reliability	- Revise the sequence of the items number. - Add the time interval in audio recording.
Practicality	Give the students more time to do the designed assessments.
Authenticity and Interactiveness	-

Backwash Effect	Provide various feedbacks.
Fairness	-

b. Trying Out, Evaluating and Revising the Second Draft of the Second Assessment Topic

1) The Try-Out

The second assessment topic was “*A Cup of Tea*”. The designed assessment was tried-out on September 18 and October 2, 2013. The first try-out followed by all the fifth grade students in SDN Rejodani. In the second try-out, 39 students joined the assessment.

2) The Evaluation

a) Students’ Attitude toward the Second Draft Assessment Designed

It is important to know how the students’ attitude toward the designed assessment. From the mean scores which are presented in the following table, it can be concluded that the students have positive attitude towards the designed assessment.

Table 77: Students’ Responses on Evaluation Questionnaires of the First Element “*A Cup of Tea*” by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
Students’ attitude toward the designed assessment	I loved the English assessment which was given.	2	3.29	Good
	I was not afraid anymore to finish the English assessment.	14	2.98	Good

Another supporting data were from the interview with the teacher and one of the students in the classroom after the first try-out of the designed assessment. The data shows that the students like the designed assessment. They can enjoy the assessment process. According the teacher, the students looked more relax when they were given the designed assessment. She had an opinion that it was caused

by the familiarity of the materials for the students. So, they were confidence to finish the designed assessment. One of the students also said that she loved the assessment because it is unique.

.....
R : ...*Kalautentangtes yang keduainibagaimana?*(Assalamualaikum, what do you think about the second designed assessment?)

T : *yang jelasanakitukelihatanlebihsantaiyambak. Mungkinkarenamaterinyasangatmerkakenal* (Students looked more relax maybe they are familiar with the materials.)

(Appendix E: The Interview Transcripts of Assessment II)

.....
R : *Eh kakakmautanyadulunihtentangtestadi. Gimanamenurutkamu?* (Well, I want to ask you about the designed assessment. What do you think?)

S : *Serusoalnyajarangada yang begitu.* (It is fun because that kind of the designed assessment is rarely to find.)

(Appendix E: The Interview Transcripts of Assessment II)

b) Relationship between the National Curriculum with the Second Draft Assessment Designed

To prove that the designed assessment is relevance with the National Curriculum, the data were taken in some ways. First, the data were obtained through the observation sheet. The observer agreed that the designed assessment is relevance with the National Curriculum.

Table 78: The Observer' Responses on Observation Sheet of the Second Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	5	It was suitable
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	6	It was suitable because the materials were the instructions.

The result of the interview with the teacher was also used as the source of the data. She had the similar opinion with the observer. The statement from the English teacher can be seen below.

.....
 R : *Apakah materinya sudah sesuai dengan SK KD dan yang diajarkan sehari-hari?* (Do you think that the materials of the designed assessment are suitable with the Standard of Competence, Basic Competence and the lesson in the classroom?)

T : *Oo sudah mbak.*

Hanya saja memang cara membuat sesuatu belum pernah saya berikan kepada anak. (Yes they are suitable. However, the steps to make something are not ever assessed yet.)

(Appendix E: The Interview Transcripts of Assessment II)

.....

The data shows the teacher agreed that the designed assessment reflected the Standard Competence, Basic of Competence and the Standard of Graduate Competence in the National Curriculum.

c) **Validity**

There are three types of the validity that should be fulfilled by the appropriate assessment. The explanation is given below.

i) **Content Validity**

The assessment tasks should be suitable with the elements in the National Curriculum and the materials which given in the classroom. Because the assessment was designed based on the students' need analysis, the product should be represented the result of the students' needs. From the excerpt below, the teacher agreed that the designed assessment is suitable with the students' needs and give some benefits to them.

.....

R : *Bagaimana dengan isi materinya bu?*

Apakah sudah sesuai dengan kebutuhan siswa? (What do you think about the materials? Do you think that it is suitable with the students' need?)

T : *Itu juga sudah kok mbak karena instruksi itu kan siswa pasti menemukan.*

Ya seperti ini misalnya akan jadi paham cara membuat sesuatu dalam bahasa Inggris. (Yes, it is because the students always find the instructions. As the example, through this designed assessment, they understand how to make something in English.)

(Appendix E: The Interview Transcripts of Assessment II)

.....

In the observation sheet, the observer as the respondent had the same opinion with the teacher. She said that the designed assessment is suitable with the

students' language competence because the level of vocabulary were easy to understand by the students.

Table 79: The Observer' Responses on Observation Sheet of the Third Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was suitable with the students' needs.	7	It was suitable.
	The designed assessment was suitable with the students' background, their ages, the culture, and their hobbies.	8	Yes
	The designed assessment was suitable with the students' language competencies.	9	The students often find the words so it was easy for them to understand.

From the mean score in the evaluation questionnaire which was administered to the students, the result shows that the designed assessment achieved the content validity. The assessment tasks were suitable with what the students have learned. It can be seen in the following table.

Table 80: Students' Responses on Evaluation Questionnaires of the Third Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment which was given was suitable with what I have learned	1	3.00	Good

ii) Construct Validity

The construction of the designed assessment should be able to assess listening skill because the aim is to assess the students' listening competencies. Both of the teacher and the observer agreed that the designed assessment is able to measure the listening skill. The data are shown in the table below.

Table 81: Respondents' Response on Observation Sheet of the Third Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	11	Yes

In the interview session with the teacher, her statement supported the data above. The excerpt can be seen as follows.

-
- R : *Apakah tes yang didesain ini sudah mampu mengukur kemampuan listening anak bu?* (Is the designed assessment able to measure the students' listening competencies?)
- T : *Kegiatannya kan mendengarkan kemudian memberikan nomor pada gambar yang tepat ya mbak jadi ya sudah sesuai menurut saya.* (The activities are listening and giving numbers to the right pictures so I think it is suitable.)
- (Appendix E: The Interview Transcripts of Assessment II)
-

iii) Face Validity

The face validity of the designed assessment can be seen through the lay-out of the assessment, the instructions, the organization of the assessment structure, etc. One of the data resources were from the evaluation questionnaire administered to the students. The mean scores for items number 4, 5 and 6 fall into good category. Those were related to the general lay-out of the designed assessment.

Table 82: Students' Responses on Evaluation Questionnaires of the Third Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	I loved the designed assessment because it was colourful.	4	3.29	Good
	I loved the designed assessment because there were a lot of pictures.	5	3.12	Good
	I could read the designed assessment easily.	6	2.83	Good

In the observation sheet, the data shows that the teacher agreed with the students. The designed assessment is good for students because it is colourful. The

worksheets also have a lot of pictures which make the students are easier to understand.

Table 83: The Teacher's Responses on Observation Sheet of the Third Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment had the interesting layout.	10	Yes because it was colourful.
	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	11	Yes
	The pictures in the designed assessment could be seen clearly.	12	Yes
	The font size in the designed assessment was readable.	13	Yes

d) Reliability

The data related to the reliability of the designed assessment were collected from two aspects. First, the reliability coefficient was obtained from assessment scores in the first and second try-out. It is 0.886 which falls into a good category. It means that the designed assessment is reliable.

Second, the data were gained through the observation sheet. Based on the statements from the teacher, the designed assessment was reliable because the number of items was suitable for children. The audio recording was also clear although it needs to be revised.

Table 84: The Teacher's Responses on Observation Sheet of the Fourth Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	14	It was suitable.
	The audio recording in the designed assessment was audible.	15	The audio recording was clear but it would be better to add sign as an indication of each part.

The students also agreed that the designed assessment was reliable. The result of the evaluation questionnaire which administered to them show that all the mean scores fall into good category.

Table 85: Students' Responses on Evaluation Questionnaires of the Fourth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the reliability of the designed assessment	The designed assessment was not too difficult to do.	3	3.02	Good
	The items of the designed assessment were not too many to do.	7	2.98	Good
	I understood the instructions of the designed assessment.	8	3.12	Good
	I could hear the sound clearly.	9	3.02	Good

From the interview session with one of the students after the first try-out, it also can be concluded that the designed assessment was reliable although the audio recording still needed to be revised. The data can be seen in the following excerpt.

-
- R : *Teruskaloulangannyabesok kayak tes yang tadimaugak?*(Do you want on the next exam, the assessment is like this?)
- S : *Mau tapijangancepet-cepet kayak tadi.*(Yes, I want. However, I do not want the audio recording is fast like this one.)
- R : *Cepet-cepet gimana? Suaranya?*(What do you mean? Do you mean the sound?)
- S : *Iya, akujadisusahkak.*(Yes. It is difficult for me.)

(Appendix E: The Interview Transcripts of Assessment II)

.....

e) Practicality

Practicality can be seen from the time allocation for students and teachers. Based on the data, the designed assessment is practical. The time allocation for students is related to the time given by the teacher for them to do the designed assessment. Based on the data which were gained from the evaluation questionnaire, the mean scores of the items number 10 and 11 are 3.07. It means

the students agreed that the time allocation for them is enough. Therefore, the designed assessment is practical.

Table 86: Students' Responses on Evaluation Questionnaires of the Fifth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The practicality of the designed assessment	I could finish all of the items in the designed assessment.	10	3.07	Good
	I could finish all of the items in the designed assessment on time.	11	3.07	Good

However, the opinion from the observer was quite different from the students. She said that the time should be extended. The data can be seen in the following table.

Table 87: The Observer' Responses on Observation Sheet of the Fifth Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the practicality of the designed assessment	Time allocation for the designed assessment was enough.	16	The time should be added more because some of the students were in hurry.

The time allocation for the teacher is related to the teacher's time frame and the scoring system. From the excerpt below, it can be concluded that the teacher has no problem related to this aspect of the practicality.

R : *Apakah ibu menemui kesulitan dengan sistem skornya?* (Do you find any difficulties related to the the scoring system?)

T : *Gak sih mbak. Biasa saja.* (No, I don't.)

R : *Apakah ibu bisa memberikan skor secara objektif and efektif dengan system ini?* (Can you give the scores objectively and effectively using the system?)

T : *Panduannya kan sudah jelas mbak, Jadi ya ndak lama kalau sudah menguasai panduannya.* (The guidance is clear so it is easy if we understand the guidance.)

(Appendix E: The Interview Transcripts of Assessment II)

f) Authenticity and Interactiveness

The data of the authenticity and the interactiveness of the second topic of the designed assessment were obtained through the evaluation questionnaire for the

students. The mean score is 3.41 which fall into a very good category. It shows that the students agreed with the statement.

Table 88: Students' Responses on Evaluation Questionnaires of the Sixth Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The authenticity and interactiveness of the designed assessment	I could find the designed assessment materials in my daily life.	12	3.41	Very Good

The data are supported by the interview sessions with the teacher. It can be concluded that the designed assessment is authentic because it reflected what the students usually find in their daily life and it gives the advantages for the students.

The result can be seen in the following excerpt.

-
- R : *Tes yang sayakembangkaninikanbermaksuduntukmenghubungkan yang biasaanaktemuidenganpelajaran, apakahinibermanfaatbu?*(The designed assessment which is designed had a purpose to connect the things which is usually found by the students with the lesson. Do you think it gave some benefits?)
- T : *Yaiya. Kalaumerekapernahmenjumpaikanjadiingintahu.*(Yes, they would want to know more if they ever found that.)
- R : *Menurutibu, apakahtesinisudahbisamenghubungkananakdengan yang merekabiasajumpai?*(In your opinion, do you think this designed assessment connect the students with the things that they usually found?)
- T : *Sudahmbak. Membuattehpastikanseringmerekalihat.*(I think so. The students must have ever seen the way to make a cup of tea.)

(Appendix E: The Interview Transcripts of Assessment II)

.....

g) Backwash Effect

A good assessment should give a good backwash effect for the students. The data can be seen in the following table.

Table 89: Students' Responses on Evaluation Questionnaires of the Seventh Element "A Cup of Tea" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The backwash effect of the designed assessment	I loved English more.	13	3.07	Good
	The designed assessment helped me to learn more about English.	15	3.12	Good

All of the items which belonged to the aspect of the backwash effect had the mean scores which fall into good category. It means that in the students' opinion, the designed assessment gave good backwash effect for them.

The teacher also had similar opinion with the students. However, she gave suggestion related to the feedback for the students. It can be seen through the result of the observation sheet which administered to them.

Table 90: The Teacher's Responses on Observation Sheet of the Seventh Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
The backwash effect of the designed assessment	The designed assessment could help the students to develop their abilities in listening skill.	18	It could help the students.
	The feedback which was given was able to improve students' competencies.	19	It would be better if the feedback was given by asking the students to make a cup of tea in front of the class.

In the following excerpt, the teacher also agreed that the designed assessment give good backwash effect for the students. It improves the students' listening skill.

R : *Apakahbermanfaatbutesini? Maksudsayapakahbisameningkatkankemampuan listening anak?*(Do you think that the designed assessment is able to improve the students' listening skill?)

T : *Jelas. Karenamerekakandilatihmbak. Padahaluntuksaatinijarangkegiatan listening untukanak.*(Yes, it is because the students are trained. Today, listening activities are rarely to find.)

(Appendix E: The Interview Transcripts of Assessment II)

h) Fairness

All of the students should get the same chances to do the designed assessment.

The facilities should be able to facilitate all of the students. The data below are taken from the observation sheet which asked the teacher related to the aspect of fairness in the designed assessment.

Table 91: The Teacher's Responses on Observation Sheet of the Eighth Element "A Cup of Tea" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
Fairness of the designed assessment	Every student got the same facilities (worksheet, audio recording, etc) in doing the designed assessment.	2	Every student got everything they need.
	Every student had the same chances to do the designed assessment.	4	OK

Based on the teacher's opinion, the designed assessment was already fair because all of the students got the same chances to do the designed assessment and all of the facilities they need. The result above is also supported by the following excerpt.

R : *Apakah menurut ibu semua anak sudah mendapat kesempatan yang sama dalam tes ini?* (Do you think that all of the students get the same chances in the designed assessment?)

T : *Kalau menurut saya sudah mbak. Semuanya sudah dapat lembar kerja, suaranya juga kedengar sampai ke belakang.* (I think so. they get the same chances because all of them got the worksheets and all of the students could hear the audio recording.)

(Appendix E: The Interview Transcripts of Assessment II)

3) The Revision

Based on the analysis of the data, the second topic of the designed assessment was already fulfilled the criteria for all the aspects of the good assessment. However, there were some parts which still needed to be revised. Each part is presented in the following table.

Table 92: The Revision of the Assessment Aspects of the Second Assessment Topic

Aspects that need to be revised	Revisions
Students' Attitude toward the Designed Assessment	-
Relationship between the National Curriculum with the Designed Assessment	-
Validity	-
Reliability	- Add the time interval in audio recording. - Add the signs in the end of each part
Practicality	Give the students more time to do the designed assessments.
Authenticity and Interactiveness	-
Backwash Effect	Give the feedback in the form of making a cup of tea in the classroom.
Fairness	-

c. Trying Out, Evaluating and Revising the Second Draft of the Third Assessment Topic

1) The Try-Out

The third assessment topic was “*Good Daily Activities*”. It was tried-out on September 21 and October 5, 2013. Forty-one students joined the first and second try-out.

2) The Evaluation

a) Students' Attitude toward the Second Draft Assessment Designed

After the first try-out, the evaluation questionnaire was distributed to the students. One of the purposes is to find out their attitude towards the designed assessment. The data can be seen in the following table.

Table 93: Students' Responses on Evaluation Questionnaires of the First Element “*Good Daily Activities*” by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
Students' attitude toward the designed assessment	I loved the English assessment which was given.	2	3.10	Good
	I was not afraid anymore to finish the English assessment.	14	2.93	Good

The mean scores for the items number two and fourteen are 3.10 and 2.93. Both of them fall into good category. Based on the data, the students had positive attitude toward the designed assessment. Another supporting data were gained through the interview session with the teacher and students. The excerpts presented as follows:

-
- R : ..., untuk yang ketiga ini kandalambentuk cerita, bagaimanamenurutibu? (..., the third designed assessment is the short story, what do you think?)
- T : Kalaumenurutsayalumayansulitmbakkarenaanakbelumpernahditessepertiini.(I think it is quite difficult for students because they are never tested like this.)
- R : Tapidari yang ibulihatselama proses tadi di kelasbagaimana? (what can you observe based on the process in the classroom?)
- T : Sayamalahtidakmenyangkajikaanaksenangmbak.
Merekajugabisamengikutidenganbaik.(I did not guess that they loved it. They could do the designed assessment well.)

(Appendix E: The Interview Transcripts of Assessment III)

-
- R : Kakakmautanyanihtentangtestadi, bolehgak? (I want to ask you about the designed assessment, is it ok?)
- S : Boleh, emangmautanyaapa? (Yes, it is. What do you want to ask?)
- R : Sukagaksamatesnya? (Did you like the designed assessment?)
- S : Akusukasoalnyaadawarnanya. (Yes, I did because it is colourful.)

(Appendix E: The Interview Transcripts of Assessment III)

.....

The teacher said that she never tested the students using short story. She guessed that it would be difficult for students. However, during the observation, the teacher found that the students enjoyed doing the third topic of the designed assessment. One of the students also said that she loved the designed assessment. It was colourful.

b) Relationship between the National Curriculum with the Second Draft Assessment Designed

The assessment designed is relevant with the National Curriculum. It was proven by the data which were got through the observation sheet. The teacher and the observer as the respondents share the same opinion as shown in the following table.

Table 94: Respondents' Response on Observation Sheet of the Second Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the correlation between the designed assessment and the National Curriculum	The designed assessment was suitable with the standard of graduate competence for elementary school in National Curriculum.	5	OK
	The designed assessment was suitable with the standard of competence and basic competence for the fifth grade students of elementary school in National Curriculum.	6	OK

From the interview session with the teacher, the data were also similar. She said that the listening assessment was already suitable with the National Curriculum. The excerpt was as follows:

R : *Jikadilihatdariisitesnya, apakahsudahsesuaibudengankurikulum yang ditentukanpemerintah?* (Based on the content, do you think it is relevant with the National Curriculum?)
T : *Sudahkokmbak. Dari SKL nyasajakansudahbisajelasdilihat.* (Yes, it is. It could be seen clearly through the Standard of Graduate Competence.)
(Appendix E: The Interview Transcripts of Assessment III)

c) Validity

The appropriate assessment should fulfil the three types of validity. They are content validity, construct validity and face validity.

i) Content Validity

The designed assessment should assess the competencies that want to be assessed. It can be seen through its relevance with the materials that the students have learned the students' needs, students' profile and the students' language competencies.

The data in the observation sheet which was administered to the teacher show that generally, the designed assessment has fulfilled the content validity criteria although there are some aspects needed to revise. The assessment tasks should

represent the activities in the students' daily life and the sentences in audio recording should be made simpler. Moreover, the observer also shared the similar ideas. The results are presented in the following tables

Table 95: The Teacher's Response on Observation Sheet of the Third Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was suitable with the students' needs.	7	It would be better if the activities were added. The examples were helping parents or doing hobbies.
	The designed assessment was suitable with the students' background, their ages, the culture, and their hobbies.	8	It was suitable.
	The designed assessment was suitable with the students' language competencies.	9	Making the sentences simpler.

Table 96: The Observer's Response on Observation Sheet of the Third Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was suitable with the students' needs.	7	The assessment tasks should be added with other activities like doing the students' hobbies
	The designed assessment was suitable with the students' background, their ages, the culture, and their hobbies.	8	It was suitable.
	The designed assessment was suitable with the students' language competencies.	9	Making the sentences simpler.

In the evaluation questionnaire, the mean score of the item which is related to the content validity is 2.85. The mean score is good which can be concluded that they agreed with the statement.

Table 97: Students' Response on Evaluation Questionnaires of the Third Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	The designed assessment which was given was suitable with what I have learned	1	2.85	Good

ii) Construct validity

Construct validity of the designed assessment demands the suitability of the assessment processes and procedures with the competencies that want to be assessed. The item number 11 in the observation sheet asks the responses from the teacher and the observer related to the construct validity. The result of the observation sheet for the teacher shows that she agreed with the statement.

Table 98: The Teacher's Response on Observation Sheet of the Third Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	11	Yes

In the interview sessions, the assessment procedures were suitable with the listening assessment activities. The excerpt is presented as follows.

-
- R : *Dari sisi prosedur pengetesan dan bentuknya, apakah sudah sesuai dengan skill yang diteskan?* (Related to the assessment form and procedures, is it suitable with the competencies that want to be assessed?)
- T : *Sudah kok. Anak-anak sudah dilatih mendengarkan.* (Yes it is. Children are trained to listen.)

(Appendix E: The Interview Transcripts of Assessment III)

.....

iii) Face Validity

Face validity deals with the general assessment appearance. Both the teacher and the observer had similar ideas related to the aspect. The data are presented in the observation sheet which is administered to them.

Table 99: Respondents' Responses on Observation Sheet of the Third Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the validity of the designed assessment	The designed assessment had the interesting layout.	10	Yes
	The designed assessment was able to measure the listening skill as what have been mentioned in the blueprint.	11	Yes
	The pictures in the designed assessment could be seen clearly.	12	Yes
	The font size in the designed assessment was readable.	13	Yes

The students share their opinion through the evaluation questionnaire. The mean scores of the items number 4, 5 and 6 are 3.07, 3.22 and 3.00. Those items represented the aspect of face validity. Since all of the mean scores fall into good category, it can be concluded that the students agreed with the statements.

Table 100: Students' Responses on Evaluation Questionnaires of the Third Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
the validity of the designed assessment	I loved the designed assessment because it was colourful.	4	3.07	Good
	I loved the designed assessment because there were a lot of pictures.	5	3.22	Good
	I could read the designed assessment easily.	6	3.00	Good

One of the students also stated that she loved the designed assessment because it was funny. The following excerpt supports the data above.

.....

R : *Suka gak sama tesnya?* (Do you like the assessment?)
 S : *Aku suka soalnya ada warnanya.* (Yes, I do because it's colourful.)
 R : *Kalo gambarnya gimana? Jelas gak?* (What do you think about the pictures? Are they clear for you?)
 S : *Udah kok, lucu juga.* (Yes, they are. They are funny too.)
 (Appendix E: The Interview Transcripts of Assessment III)

.....

d) Reliability

The data which supported the reliability of the designed assessment were got through the analysis of the data of the first try-out and the result of the test-retest analysis of the first and the second assessment scores.

The result of the analysis of the assessment scores shows that the designed assessment is reliable since the Alpha's score is 0.895. The result of the observation sheet shows the number of items in the designed assessment was already suitable for children. However, time interval in audio recording needed to be lengthened. The purpose was making the students easier to understand and giving more time for them to think.

Table 101: The Observer's Response on Observation Sheet of the Fourth Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the reliability of the designed assessment	The designed assessment had the number of items which were suitable for children.	14	It was suitable.
	The audio recording in the designed assessment was audible.	15	The audio recording was good but it would be better to make it lengthen.

Both of the students and teacher had the similar opinion about the audio recording. They agreed that it needed to be revised as shown in the following excerpts.

.....
 R : *Oiya, tadikantesnyamendengarkan, suaranya jelas gak?* (It is the listening assessment. Do you think that the sounds is clear?)

S : *Jelastapikecepatan. Belumslesai, udahngomonglagi.* (It is clear but too fast. I did not finish yet but the next number started.)

(Appendix E: The Interview Transcripts of Assessment III)

.....
 R : *.....Bagaimanadengansuararekamannyabu?Apakahcukupbaik?* (How is the audio recording? Do you think that it is good?)

T : *Sudahkokmbakhanyasajamungkinkalimatnyabisalebihsederhanasupayalebihmud*

ahdipahamidandiberikanjedalebih lama untukanak. (Yes, it is but maybe the sentences should be made simpler and the audio recording is added more time interval.)

(Appendix E: The Interview Transcripts of Assessment III)

e) Practicality

The designed assessment should be feasible to do. The allocation time for students to do the assessment and the time which needed by the teacher in scoring system are the factors. The third topic of the assessment already fulfilled those criteria. It was proven by the following data.

Table 102: Respondents' Responses on Observation Sheet of the Fifth Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
the practicality of the designed assessment	Time allocation for the designed assessment was enough.	16	OK

The data above were gained from the observer. The following excerpt supported the data above. It was the interview session with one of the students.

R : *waktunyacukupgakbuatngerjainsemuasalnya?* (Is the time enough to finish all of the items?)

S : *cukupkok.* (Yes, it is.)

(Appendix E: The Interview Transcripts of Assessment III)

The result of evaluation questionnaire for students is similar with the data above. The mean scores are 2.83 and 2.90. Both of them fall into good category which showed that the third topic of the assessment was practical.

Table 103: Students' Responses on Evaluation Questionnaires of the Fifth Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The practicality of the designed assessment	I could finish all of the items in the designed assessment.	10	2.83	Good
	I could finish all of the items in the designed assessment on time.	11	2.90	Good

Related to the time allocation for teacher, she said that the scoring system is still manageable. She can use it effectively.

R : *Berkaitan dengan panduan penilaiannya, apakah ada kesulitan bu untuk menggunakannya?* (Related to the scoring guidance, is it difficult for you to use it?)

T : *Menurut saya itu efektif kok mbak. Itu kriterianya diambil berdasar teori kan ya?* (I think it's effective. Did the criteria are based on the theory?)

R : *Oh iya ibu. Kriterianya sesuai dengan apa yang harus dilihat dari kemampuan listening anak.* (Yes, it is. The criteria made based on the listening competencies for children.)

T : *Yah kalo begitu, berarti sudah tidak ada masalah mbak.* (So, I think there is no problem about it.)

(Appendix E: The Interview Transcripts of Assessment III)

f) Authenticity and Interactiveness

Related to the authenticity and the interactiveness of the designed assessment, it is proven by the data from the observation sheet. Based on the data in the observation sheet which was administered to the teacher on the first try-out, the designed assessment is authentic and interactive because the assessment tasks are relevant with the students' daily life.

Table 104: Respondents' Responses on Observation Sheet of the Sixth Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
The authenticity and interactiveness of the designed assessment	The designed assessment was suitable with students' daily life.	17	It was suitable because the materials were about the students' activities in their daily life.

The result of the evaluation questionnaire is also supported the data above.

The mean score for the item is 2.83 which falls into a good category. It means that the students agree with the statement.

Table 105: Students' Responses on Evaluation Questionnaires of the Sixth Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The authenticity and interactiveness of the designed assessment	I could find the designed assessment materials in my daily life.	12	2.83	Good

g) Backwash Effect

It was important to find out whether the designed assessment gave good backwash effect to the students or not. It can be indicated through the students' motivation and the influence of the deigned assessment through their learning process.

Table 106: Students' Responses on Evaluation Questionnaires of the Seventh Element "Good Daily Activities" by Question Number, Mean Scores and Category

Elements of the Items	Items	Question Number	Mean Scores	Category
The backwash effect of the designed assessment	I loved English more.	13	3.17	Good
	The designed assessment helped me to learn more about English.	15	2.98	Good

From the table above, it can be seen that the mean scores for all of the items in the seventh element were good. It shows the listening assessment gave a good back wash effect for the students. From the excerpt below, it was also proven that the designed assessment gave good backwash effect to the students.

.....
R : *Apakah anak-anak menyuka bahasa Inggris?* (Do the students love English?)

T : *Dari yang saya lihat itu sudah bakti sulit menurut mereka.* (Based on my observation, they love it but it is difficult for them.)

R : *Menurut ibu, apakah tes ini dapat memotivasi mereka untuk belajar lebih rajin lagi dan membuat anak-anak lebih menyuka bahasa Inggris?* (In your opinion, do you think the designed assessment can motivate the students to learn more and to love English?)

T : *Menurut saya bisa karena materinya menarik dan tidak seperti biasanya yang mereka temukan.* (Yes, it is. The materials are interesting and different from the materials that the students are usually found.)

Appendix E: The Interview Transcripts of Assessment III)

.....

h) Fairness

All of the students should get the same chances in doing the designed assessment. Both the teacher and the observer stated their agreement in the observation sheet. The data shows that they had similar opinion as listed in the following table.

Table 107: Respondents' Responses on Observation Sheet of the Eighth Element "Good Daily Activities" by Question Number and Statements

Elements of the Items	Items	Question Number	Statements
Fairness of the designed assessment	Every student got the same facilities (worksheet, audio recording, etc) in doing the designed assessment.	2	Yes
	Every student had the same chances to do the designed assessment.	4	Yes

Based on the result in the table 107, the designed assessment achieved the aspect of fairness because all of the students got the same facilities and everything they need during the process.

In the interview sessions with one of the students, she said that she could hear the audio recording clearly although she sat on the back. All of the students also got the worksheets.

R :*menurutkamu tesnya adil gak?* (Do you think the designed assessment fair?)

S : *Adil gimana maksudnya?* (What do you mean?)

R : *Maksudnya apa semua murid bisa dengar suara rekamannya? Kankamu ada dibelakang, kedengar gak?* (did all of the students can hear the audio recording? You sat in the last row. Could you hear the recording?)

S : *Kedengar kok. Menurutku si hadil.* (Yes, I could. I think it is fair.)

R : *Lembartesnya kan semua jugadapat ya.* (All of the students also get the worksheet, aren't they?)

S : *Iya.* (Yes, they are.)

Appendix E: The Interview Transcripts of Assessment III)

3) The Revision

Based on the analysis of the try-out of the third assessment topic, generally the designed assessment already fulfilled the criteria for all of the aspects. However, there were some parts which needed to be improved. The revisions are presented in the following table.

Table 108: The Revision of the Assessment Aspects of the Third Assessment Topic

Aspects that need to be revised	Revisions
Students' Attitude toward the Designed Assessment	-
Relationship between the National Curriculum with the Designed Assessment	-
Validity	<ul style="list-style-type: none"> - Add the activities like helping parents or doing hobbies - Make the sentences simpler.
Reliability	Add the time interval in audio recording
Practicality	-
Authenticity and Interactiveness	-
Backwash Effect	-
Fairness	-

4. The Final Draft of the Listening Assessment

The second draft had been tried-out. During the try-out, the teacher and observer gave their opinions and suggestions for each topic. The evaluation questionnaire was also distributed to the students to get their opinions and suggestions also. There were also the interview sessions with the teacher and the students after the try-out. All of the data above was used to revise the second draft.

The final draft was made based on all of the data which were obtained through all of the instruments made. The final draft of the designed listening assessment consisted of the three topics which made based on the learners' need analysis, the National Curriculum, the opinions and suggestions from the students and the

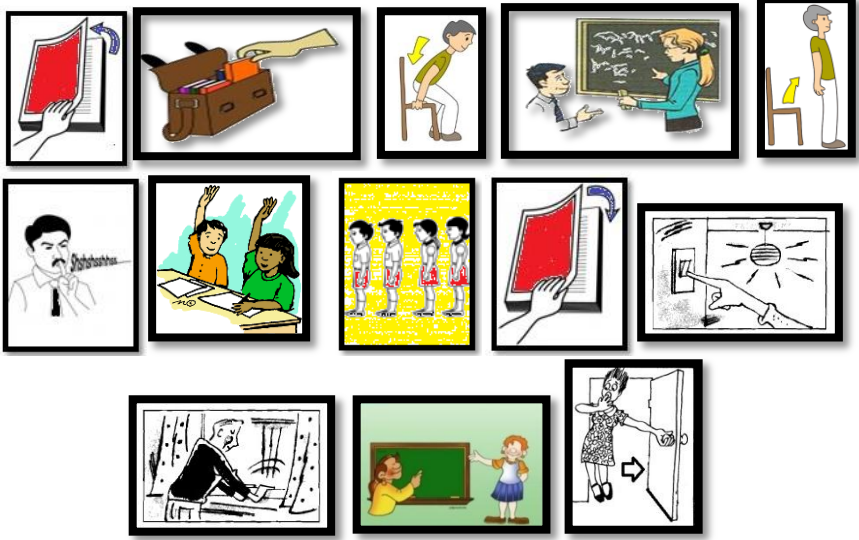
respondents. The specifications of the final draft listening assessment designed for each topic can be seen below.

a. First Assessment (*In the Classroom*)

Finally, the first topic of the assessment consisted of ten items. The sequence of numbers was also already revised. The time interval of audio recording and the allocation time for students had been added more and the pictures were made bigger. The detail of the specification can be seen in the following table.

Table 109: The Final Draft of the First Assessment Specifications

Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungansekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah
Skill	Listening - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macroskill</i>) - Retain chunks of language of different lengths in short-term memory (<i>microskill</i>)
Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.
Assessment Objective(s)	Assessing students' ability to identify the simple classroom instructions by labelling the correct pictures based on the audio recording about the classroom instructions.
Text Types	The topic was <i>In the Classroom</i> . The assessment materials were simple instructions which were usually done in the classroom and it would be presented in recorded form.
Main Sub-Section of the Text	The students were given \pm 30 minutes to finish one part of the designed assessment which consisted of ten items.
Language Competence	- vocabulary: topical vocabulary related to the classroom instructions - syntax: simple sentences and imperatives
The Input	- Language (Audio Recording) <div style="display: flex; justify-content: space-between;"> <div> Sit Down. (example) Be Quiet. Clean the board Raise your hand. Close your book. Stand up. </div> <div> Make a line. Open the door. Turn on the lamp. Close the window. Open the book </div> </div>





	<p>- Non-Language (Pictures)</p> 
The Expected Responses	non-language (labelling correct picture by giving a tick (√))
Criteria of Marking	<ul style="list-style-type: none"> - Each of the correct answer got 1 point - Each of the wrong answer got 0



b. Second Assessment (*A Cup of Tea*)

The second topic of the assessment had been revised in the aspects of reliability, practicality, and the backwash effect. The time interval in audio recording and the allocation time for students had been added more. The signs in the end of each part in audio recording also had been added. To enhance the good backwash effect, the instructions on giving feedback in the teacher's guide also had been revised. The detail of the specification can be seen in the following table.

Table 110: The Final Draft of the Second Assessment Specifications

Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah

Skill	<p>Listening</p> <ul style="list-style-type: none"> - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macroskill</i>) - Retain chunks of language of different lengths in short-term memory (<i>microskill</i>) 						
Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.						
Assessment Objective(s)	<ol style="list-style-type: none"> 1. Assessing students' ability to number the right pictures based on the materials and ingredients which are needed to make a cup of tea in audio recording. 2. Assessing students' ability to identify simple imperative statements by arranging and giving numbers to the pictures based on the steps of making a cup of tea in the audio recording. 						
Text Types	The topic was <i>A Cup of Tea</i> . The assessment materials were materials and simple instructions which were related to the ways of making a cup of tea. It would be presented in recorded form.						
Main Sub-Section of the Text	The students were given \pm 35 minutes to finish two parts of the designed assessment. Part A had five items while part B had seven items.						
Language Competence	<ul style="list-style-type: none"> - vocabulary: topical vocabulary related to the materials and steps to make a cup of tea - syntax: simple sentences and imperatives 						
The Input	<p>- Language (Audio Recording)</p> <p>Part A</p> <table border="0"> <tr> <td>some water (example)</td><td>some milk</td></tr> <tr> <td>a cup and a spoon</td><td>some sugar</td></tr> <tr> <td>a teabag</td><td>a kettle</td></tr> </table> <p>Part B</p> <ol style="list-style-type: none"> 1. First, put some water in the kettle. (example) 2. Second, boil the water. 3. Third, add some sugar in the cup. 4. Fourth, put a tea bag in the cup. 5. Fifth, pour the hot water into the cup carefully. 6. Sixth, add some milk in the cup. 7. Seventh, stir it firmly. 8. Finally, we get a cup of delicious tea. <p>- Non-Language (Pictures)</p> <p>Part A</p> <div style="display: flex; justify-content: space-around; align-items: center;">     </div>	some water (example)	some milk	a cup and a spoon	some sugar	a teabag	a kettle
some water (example)	some milk						
a cup and a spoon	some sugar						
a teabag	a kettle						


	 <p>Part B</p> 
The Expected Responses	<p>Part A: non-language (giving numbers to the correct pictures)</p> <p>Part B: non-language (arrange the pictures by giving numbers)</p>
Criteria of Marking	<p>Part A</p> <ul style="list-style-type: none"> - Each of the correct answer got 2 points - Each of the wrong answer got 0 <p>Part B</p> <ul style="list-style-type: none"> - Each of the correct answer got 1 point - Each of the wrong answer got 0

c. Third Assessment (*Good Daily Activities*)

The third topic of the assessment consisted of eleven items. The sentences were made simpler and the materials were added by the activities of doing the hobby. The time interval of audio recording also had been added more. The detail of the specification can be seen in the following table.

Table 111: The Final Draft of the Third Assessment Specifications

Standard of Graduate Competence	Memahami instruksi, informasi dan cerita sangat sederhana yang disampaikan secara lisan dalam konteks kelas, sekolah dan lingkungan sekitar
Standard of Competence	5. Memahami instruksi sangat sederhana dengan tindakan dalam konteks sekolah
Basic of Competence	5.1. Merespon instruksi sangat sederhana dengan tindakan secara berterima dalam konteks sekolah
Skill	<p>Listening</p> <ul style="list-style-type: none"> - Recognizing the communicative functions of utterances according to the situations, participants and goals (<i>macro skill</i>) - Retain chunks of language of different lengths in short-term memory (<i>micro skill</i>)

Participants	Fifth grade students of elementary school which were around 10-11 years old. They come from different social and economic background. Most of their parents worked as farmers, teachers and employees.
Assessment Objective(s)	Assessing students' ability to identify simple good daily activities by putting the number on the right picture based on the story in the audio recording.
Text Types	The topic was <i>Good Daily Activities</i> . The assessment materials were several daily activities which were usually done and it would be presented in recorded form.
Main Sub-Section of the Text	The students were given \pm 25 minutes to finish one part of the designed assessment which consisted of eleven items.
Language Competence	<ul style="list-style-type: none"> - vocabulary: topical vocabulary related to good daily activities - syntax: simple sentences
The Input	<p>- Language (Audio Recording) I get up at five o'clock in the morning. Then, I take a bath and get dressed. I have breakfast before I go to school. I brush my teeth after having it. Then, I go to school. At the break time, I have my lunch. I go home at twelve o'clock. I have a break then I ride my bicycle with my brother. After that, I do my homework. At seven o'clock, I have dinner. I enjoy my evening by watching TV with my family. At nine o'clock, I go to bed.</p> <p>- Non-Language (Pictures)</p> 
The Expected Responses	non-language (arrange the pictures by giving numbers)
Criteria of Marking	<ul style="list-style-type: none"> - Each of the correct answer got 5 points - Each of the wrong answer got 0

K. Discussion

Based on the data analysis in the previous part of the chapter, it can be concluded that the designed assessment is able to accommodate the students' needs and suitable with their characteristics as children. Moreover, the designed assessment has fulfilled the criteria of the appropriate assessment.

The result of need analysis shows that the students love English and want to learn English. However, they are afraid of getting bad mark in the listening assessment which is made by the teacher. In other aspects, they also state that they love the assessment which has pictures and colourful.

The appropriate listening assessment should fulfil some of the criteria which are validity, reliability, practicality, having good backwash effect, fairness, authenticity and interactiveness and relevance with the National Curriculum. The further discussions related to the result of the data analysis about those criteria can be seen below.

a. Validity

Based on the data analysis related to the validity of the assessment, all of the three topics have fulfilled the validity of the assessment. In the first assessment, to enhance the validity, the worksheet is made bigger. The second assessment is valid without revision. The third assessment is valid. However, there are improvements related to the sentences and the assessment tasks of the assessment.

b. Reliability

Reliability of the assessment can be seen from some views. Related to the audio recording, the sounds are clear and audible. However, the interval time in the recording is needed to be added more. The designed listening assessment also

provides the scoring guidelines to avoid bias from the assessor. The test-retest method which used to assess the consistency of measurement also shows that the designed assessment is reliable.

c. Practicality

In terms of the practicality, all of the respondents agree that the students need time more to do the designed listening assessment. Based on the result of the interview with the teacher, the reason is because the assessment is quite new for the students. Related to the time allocation in the scoring system, the teacher has no problems.

d. Good Backwash Effect

The students agree that the designed listening assessment can motivate them to learn English and they are not afraid anymore to face the listening assessment. From the teacher and the observer's view, the assessment also gives the good effect to the students. However, the second assessment can give more the good effect if the feedback is in practical experience.

e. Fairness

Related to the criteria of fairness, the designed listening assessment gives the same chances to all of the students to do the assessment. It includes the worksheets and the audio recording in the assessment process. The supporting data can be seen in the previous part.

f. Authenticity and Interactiveness

All of the respondents agree that the designed listening assessment is authentic and interactive. The materials are related to the students' daily life. In the first assessment, the materials are classroom instructions which are given everyday to

the students in the classroom. The second assessment's materials focus on simple instructions of making a cup of tea which can be found everyday by the students in their lives. The materials of the third assessment are the daily activities which are usually done by the students.

g. Relationship between the Designed Listening Assessment and the National Curriculum

All of the assessments are made based on the National Curriculum. It is proven by the result of the data analysis which was got from the experts, the teacher and the observer.

CHAPTER V

CONCLUSIONS, IMPLICATIONS AND SUGGESTIONS

A. Conclusions

The aim of the research study is to design a good listening assessment which is appropriate for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta. Based on the result of the research study, the product is successfully designed. The reasons can be seen as follows.

1. The Designed Listening Assessment for the Fifth Grade Students of SDN Rejodani Ngaglik Sleman Yogyakarta

The listening assessment is designed in three topics. They are *In the Classroom*, *A Cup of Tea* and *Good Daily Activities*. All of the topics are made based on the students' need and characteristics. The materials are related to the National Curriculum. The worksheet for each topic is designed with a lot of colourful pictures and simple instructions. The recordings are adjusted with the students in terms of the speed, the pitch and the time interval.

The information above answers the first research question which had already stated in chapter I related to how the designed listening assessment for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta should be looked like.

2. The Characteristics of the Appropriate Listening Assessment Tasks for the Fifth Grade Students of SDN Rejodani Ngaglik Sleman Yogyakarta

Based on the research study, there are some important characteristics of the appropriate listening assessment design for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta. An appropriate listening assessment should be valid, reliable, practical, having good backwash effect for students, fair, authentic and interactive. Moreover, the materials which were assessed should be based on the National Curriculum.

a. Validity

The designed listening assessment was valid since it assessed that should be assessed which was listening skill. The layout of the worksheet was already fulfilled the criteria of face validity.

b. Reliability

The designed listening assessment for the fifth grade students of SDN Rejodani was reliable. It can be seen from two aspects. First, the designed assessment provided scoring guidelines. It would make the assessor was easier to interpret the result accurately and assigned the suitable assessment scores to the students. Second, the reliability can be seen from the physical content of the designed assessment itself like every student got the worksheet and the audio recording was audible for all of the students.

c. Practicality

The product was set into the specific time for each topic depended on the difficulty level of the each assessment topic. Therefore, based on the result of the research study, the designed listening assessment was practical.

d. Good Backwash Effect

The product was able to motivate the students in learning English more. They were also comfortable and were not afraid to face the listening assessment. It means that the designed listening assessment gave good backwash effect for the students.

e. Fairness

All of the students got the same chances during the assessment process. The worksheets of the designed listening assessment were already distributed for all of the students. The audio recording was clear for all of the students. The scoring guidelines which were provided for each topic was able to avoid the bias in assess the students' competencies.

f. Authenticity and Interactiveness

The authenticity and the interactiveness of the designed assessment can be seen from some aspects. First, the materials were suitable with the students' daily life. Second, the topics were meaningful and the language of the assessment was natural.

g. Relation between the Designed Assessment and the National Curriculum

The designed listening assessment for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta was designed based on the Standard of

Graduate Competence, the Standard of Competence and Basic Competencies in National Curriculum. Therefore, the product was suitable with the National Curriculum. It was supported by the data from expert judgements and the data during the try-out of the product.

The information above answers the second research question which had already stated in chapter I related to the characteristics of the appropriate assessment tasks for the fifth grade students of SDN Rejodani Ngaglik Sleman Yogyakarta.

B. Implications

There are some implications related to the conclusions above. Those can be seen in the following paragraphs.

First, the listening assessment for the fifth grade students of SDN Rejodani is designed based on the students' needs and the National Curriculum. The model is also adjusted to make it suitable with the children characteristics as the test-takers. The implication from the aspect is that the designed listening assessment as the product of the research can be used in other elementary schools.

Second, a good listening assessment should have some characteristics like what have been mentioned above. Based on the result of the research study, the listening assessment which was designed has already fulfilled those characteristics. The data showed that the listening assessment which fulfilled those characteristics is suitable with the students' characteristics and able to assess the students' competencies accurately. Therefore, if the English teacher wants to

make the listening assessment for the students, she should make the listening assessment which has those criteria.

C. Suggestions

1. The English Teachers

It has been proven that the suitable and appropriate listening assessment can motivate the students to learn more and is able to measure the students' competencies in listening skill accurately. The English teacher should be able to create the appropriate listening assessment for the students.

2. The Principals of the Elementary Schools

As the principals, it is important to control all of the teachers in the school. It is needed to check whether the process of teaching and learning is effective or not. Related to the assessment process, the principals are expected to support the teacher. It can be in the form of facilities, funds, etc. It is really needed to understand that assessing the students is not only about the marks but also about their improvements of knowledge.

3. English Education Department

The result of the research study is not only increasing the number of collections but also to inspire other researchers. Moreover, it is expected to give the students of English Education Department the motivation to know more about the language assessment for children because it has the own specifications which is different from the assessment for adults

4. The Government

It is expected that the government give more attention in the assessment process in Elementary School because somehow it is not suitable for the students. The ways, techniques, methods and facilities in assessing students in Elementary School are different from other school levels because of the age level of students.

5. Other Researchers

The research study is expected to encourage other researchers to conduct the similar research with different skills and group of learners. It is also possible to design other techniques to assess listening skill for young learners.

REFERENCES

- Bachman, L. F., & Palmer, A. S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Hong Kong: Oxford University Press
- Baxter, Andy. 1997. *Evaluating Your Students*. London: Richmond Publishing
- Borg, W. R., & Gall, M. D. 1983. *Educational Research and Development* (4th Edition ed.). New York: Longman Inc.
- Brewster, J., Ellis, G., & Gurrard, D. 2002. *The Primary English Teacher's Guide* (New Edition ed.). Edinburgh: Pearson Education Limited
- Brown, H. D. 2004. *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education, Inc.
- Buck, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press.
- Cameron, Lynne. 2001. *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press
- Committee on Developmental Outcomes and Assessments for Young Children. 2008. *Early Childhood Assessment: Why, What and How*. Washington DC: The National Academies Press
- Departemen Pendidikan Nasional. 2008. *Pedoman Pembelajaran Bahasa Inggris di Sekolah Dasar*. Jakarta: Depdiknas
- Ferguson, M., Diptoadi, V. L., Halimi, S. S., Herini, M. P., Kismadi, G. C., Kusuma, P. C., et al. (2009). *Becoming A Creative Teacher: A Manual for Teaching English to Indonesian Elementary Students*. Jakarta, Indonesia.
- Freeman, Y., and Freeman, D. 2003. "Struggling English Language Learners: Keys for Academic Success". *TESOL Journal*, 3, XIII, page.18-23

- Garcia, E. E., Lawton, K., & Diniz de Figueiredo, E. H. 2010. Assessment of Young English Language Learners in Arizona. *The Civil Rights Project* .
- Georgiou, S. I., & Pavlou, P. 2003. *Assesing Young Learners*. (A. Maley, Ed.) New York: Oxford University Press.
- Gordon, T. 2007. *Teaching Young Children a Second Language*. (D. P. Fromberg, & L. R. Williams, Eds.) Westport: Praeger Publisher.
- Harris, Michael., andMc Cann, Paul. 1994. *Handbooks for the English Classroom: Assessment*. Oxford: Macmillian Publishers Ltd
- Linse, C. T. 2005. *Practical English Language Teaching: Young Learners*. (D. Nunan, Ed.) New York: Mc Graw-Hill.
- Masuhara, H. 1998. What do teachers really want from coursebooks? In B. Tomlinson, *Material Development in Language Teaching*. Cambridge: Cambridge University Press.
- McKay, Penny. 2006. *Assessing Young Language Learners*. Cambridge: Cambridge University Press
- Nation, I. S. P., and Newton, Jonathan. 2009. *Teaching ESL/ EFL Listening and Speaking*. New York: Routledge Taylor & Francis Group
- Nunan, David. 2004. *Task-Based Language Teaching*. New York: Cambridge University Press
- Ontario Ministry of Education. 2011. *The Full-Day Early-Learning Kindergarten Program*. Ontario: Ontario Ministry of Education
- Pinter, A. 2006. *Teaching Young Language Learners*. China: Oxford University Press.
- Richards, J. C., & Schmidt, R. 2002. *Dictionary of Language Teaching and Applied Linguistics* (3rd Edition ed.). Edinburgh: Longman.

- Rivers, W. M. 1981. *Teaching Foreign-Language Skills* (2nd Edition ed.). Chicago: The University Chicago Press.
- Shaaban, Kassim. 2001. "Assessment of Young Learners". *FORUM Journal*, 4,152,XXXVIII, page. 16-17
- Scott, W., & Ytreberg, L. 1990. *Teaching English to Children*. New York: Longman Inc.
- Slattery, M., & Willis, J. 2001. *English for Primary Teachers*. Oxford: Oxford University Press.
- University of Cambridge ESOL Examinations. 2009. *Cambridge Flyers 6 Answer Booklet*. Cambridge: Cambridge University Press
- Wakeford, R. 2003. Principles of Student Assessment. In *A Handbook for Teaching and Learning in Higher Education* (p. 42). London: Kogan Page Limited.